



## Fundamental Study

# The recognizability of sets of graphs is a robust property<sup>☆</sup>

Bruno Courcelle<sup>a,\*</sup>, Pascal Weil<sup>b</sup>

<sup>a</sup>*LaBRI, Université Bordeaux-1, 351 cours de la Libération, 33405 Talence Cedex, France*

<sup>b</sup>*LaBRI, CNRS, 351 cours de la Libération, 33405 Talence Cedex, France*

Received 18 January 2005; accepted 9 March 2005

Communicated by M. Nielsen

### Abstract

Once the set of finite graphs is equipped with an algebra structure (arising from the definition of operations that generalize the concatenation of words), one can define the notion of a recognizable set of graphs in terms of finite congruences. Applications to the construction of efficient algorithms and to the theory of context-free sets of graphs follow naturally. The class of recognizable sets depends on the signature of graph operations. We consider three signatures related respectively to Hyperedge Replacement (HR) context-free graph grammars, to Vertex Replacement (VR) context-free graph grammars, and to modular decompositions of graphs. We compare the corresponding classes of recognizable sets. We show that they are robust in the sense that many variants of each signature (where in particular operations are defined by quantifier-free formulas, a quite flexible framework) yield the same notions of recognizability. We prove that for graphs without large complete bipartite subgraphs, HR-recognizability and VR-recognizability coincide. The same combinatorial condition equates HR-context-free and VR-context-free sets of graphs. Inasmuch as possible, results are formulated in the more general framework of relational structures.

© 2005 Elsevier B.V. All rights reserved.

**Keywords:** Recognizable set of graphs; Graph algebra; Hyperedge replacement; Vertex replacement; Quantifier-free definable operation; Locally finite congruence; Modular decomposition

<sup>☆</sup> This paper was prepared in part while the second author was an invited Professor at the University of Nebraska-Lincoln. The second author acknowledges partial support from the AS 93 of the Département STIC of CNRS, and the ACI Sécurité Informatique of the Ministère de la Recherche.

\* Corresponding author.

E-mail addresses: [bruno.courcelle@labri.fr](mailto:bruno.courcelle@labri.fr) (B. Courcelle), [pascal.weil@labri.fr](mailto:pascal.weil@labri.fr) (P. Weil).

## 1. Introduction

The notion of a recognizable language is a fundamental concept in Formal Language Theory, which has been clearly identified since the 1950s. It is important because of its numerous applications, in particular for the construction of compilers, and also for the development of the Theory: indeed, these languages can be specified in several very different ways, by means of *automata*, *congruences*, *regular expressions* and *logical formulas*. This multiplicity of quite different definitions is a clear indication that the notion is central since one arrives at it in a natural way from different approaches. The equivalence of definitions is proved in fundamental results by Kleene, Myhill and Nerode, Elgot and Büchi.

The notion of a recognizable set has been extended in the 1960s to trees (actually to trees representing finite algebraic terms), to infinite words and to infinite trees. In the present article we discuss its extension to sets of finite graphs.

The recognizability of a set of finite words or trees can be defined in several ways, as mentioned above, and in particular by finite *deterministic* automata. This definition (together with the related effective translations from other definitions) provides linear-time recognition algorithms, which are essential for compiler construction, coding, text processing, and in other situations. Recognizable sets of words can also be defined in an algebraic way by finite saturating congruences relative to the monoid structure. These definitions, by automata and congruences, extend smoothly to the case of finite trees (i.e., algebraic terms), using the natural algebra structure. The notion of recognizability in a general algebra is due to Mezei and Wright [37]. We will not discuss here the extensions to infinite words and trees, which raise specific problems surveyed by Thomas [43] and Perrin and Pin [40]. Our aim will be to consider sets of finite graphs.

For finite graphs, there is no automaton model, except in very special cases, and in particular in the case of graphs representing certain *labelled partially ordered sets* and *traces* (a trace is a directed acyclic graph, representing the equivalence class of a word w.r.t. a partial commutation relation), see the volume edited by Diekert [22] and the papers by Lodaya and Weil [32,33] and Ésik and Németh [24]. Algebraic definitions via finite congruences can be given because the set of finite graphs can be equipped with an algebraic structure, based on graph operations like the concatenation of words. However, many operations on graphs can be defined, and there is no prominent choice for a standard algebraic structure like in the case of words where a unique associative binary operation is sufficient. Several algebraic structures on graphs can be defined, and distinct notions of recognizability follow from these possible choices. It appears nevertheless that two graph algebras, called the *HR-algebra* and the *VR-algebra* for reasons explained below, emerge and provide robust notions of recognizability. The main purpose of this paper is to demonstrate the robustness of these notions. By robustness, we mean that taking variants of the basic definitions does not modify the corresponding classes of recognizable sets of graphs.

In any algebra, one can define two family of sets, the recognizable sets and the equational sets. The equational sets are defined as the components of the least solutions of certain systems of recursive set equations, written with set union and the operations of the algebra, extended to sets in the standard way. Equational sets can be considered as the natural extension of context-free languages in a general algebraic framework (Mezei and Wright [37], Courcelle [13] for a thorough development). The two graph algebras introduced above,

the HR- and the VR-*algebra*, are familiar to readers interested in graph grammars, because their equational sets are the (context-free) *Hyperedge Replacement* (HR) sets of graphs on the one hand, and the (context-free) *Vertex Replacement* (VR) sets on the other. Both classes of context-free sets of graphs can be defined in alternative, more complicated ways in terms of graph rewritings, and are robust in the sense that they are closed under certain transformations expressible in Monadic Second-Order Logic [15].

The main results of this paper, described below in more detail, are:

- (1) the robustness of the classes of VR- and HR-recognizable sets of graphs,
- (2) the robustness of the class of recognizable sets of finite relational structures (equivalently of simple directed ranked hypergraphs), which extends the two previous classes,
- (3) the exhibition of structural conditions on sets of graphs implying that HR-recognizability and VR-recognizability coincide,
- (4) the comparison of the recognizable sets of the VR-algebra and those of a closely related algebra representing *modular decompositions* (modular decomposition is another useful notion for graph algorithms).

The notion of recognizability of a set of finite graphs is important for several reasons. First, because recognizability yields linear-time algorithms for the verification of a wide class of graph properties on graphs belonging to certain finitely generated graph algebras. These classes consist of graphs of bounded tree-width and of bounded clique-width. These two notions of graph complexity are important for constructions of polynomial graph algorithms, see [23,20]. Furthermore, these graph properties are not very difficult to identify because Monadic second-order (MS) logic can specify them in a formalized and uniform way. (In many cases, an MS formula can be obtained from the graph theoretical expression of a property.) More precisely, a central result [8,9,15,20] says that every set of graphs (or graph property) definable by an MS formula is recognizable (respectively admits such algorithms), for appropriate graph algebras. This general statement covers actually several distinct situations.

Another reason comes from the theory of Graph Grammars. The intersection of a context-free set of graphs and of a recognizable set is context-free (in the appropriate algebraic framework). This gives immediately many closure properties for context-free sets of graphs, via the use of MS logic as a specification language for graph properties. Recognizability also makes it possible to construct terminating and (in a certain sense) confluent graph rewriting rules by which one can recognize sets of graphs of bounded tree-width by graph reduction in linear time, see Arnborg et al. [2].

Finally, recognizability is a basic notion for dealing with languages and sets of terms, and on this ground, its extension to sets of graphs is worth investigating. Logical characterizations of recognizability can be given using MS logic, extending many results in language theory [16,24,28–30]. Several questions remain open in this research field.

We have noted above that defining recognizability for sets of graphs cannot be done in terms of finite automata, so that the algebraic definition in terms of finite congruences has no alternative. Another advantage of the algebraic definition is that it is given at the level of universal algebra [37], and thus applies to objects other than graphs. However, even in the case of graphs, the algebraic setting is useful because it hides (temporarily) the complexities of operations on graphs and makes it possible to understand what is going on at a structural level.

We now present the main results of this article more in detail. The two main algebraic structures on graphs called VR and HR, originate from algebraic descriptions of context-free graph grammars. Definitions will be given in the body of the text. It is enough for this introduction to retain that the operations of VR are more powerful than those of HR. Hence every HR-context-free set of graphs (i.e., defined by a grammar based on the operations of HR) is VR-context-free, but not vice-versa. For recognizability, the inclusion goes in the opposite direction: every VR-recognizable set is HR-recognizable but the converse is not true. However, if the graphs of a set  $L$  have no subgraph of the form  $K_{n,n}$  (the complete bipartite graph on  $n + n$  vertices) for some  $n$ , then  $L$  is HR-recognizable if and only if it is VR-recognizable (this is the main theorem of Section 6). A similar statement is known to hold under the same hypothesis for context-free sets: if  $L$  is without  $K_{n,n}$  (i.e., no graph in  $L$  contains a subgraph isomorphic to  $K_{n,n}$ ), then it is HR-context-free if and only if it is VR-context-free [12]. The proofs of the two statements are however different (and both difficult).

Up to now we have only discussed graphs, but our approach, which extends the approach developed by Courcelle in [9], also works for hypergraphs and for relational structures.

The operations on graphs, hypergraphs and structures are basically of three types defined in Section 3: we use only one binary operation, the disjoint union; we use unary operations defined by quantifier-free first-order formulas; and basic graphs and structures corresponding to nullary operations. In this way we can generate graphs and structures by finite algebraic terms. The quantifier-free definable operations can modify vertex and edge labels, add or delete edges. This notion is thus quite flexible. What is remarkable is that these numerous operations can be added without altering the notion of recognizability.

The main result of Section 4 states that the same recognizable sets of graphs are obtained if one uses the basic VR-algebra (closely connected to the definition of *clique-width*), the same algebra enriched with quantifier-free definable operations, and even the larger algebra dealing with relational structures. Variants of the VR-algebra which are useful, in particular for algorithmic applications, are also considered, and they are proved to yield the same class of recognizable sets.

In Section 5, we discuss similarly the HR-algebra which is very important because of its relation with *tree-width* and with context-free graph grammars. We prove a robustness result relative to the subclass such that the distinguished vertices denoted by distinct labels (nullary operations) are different. The HR-operations are appropriate to handle graphs and hypergraphs with multiple edges and hyperedges (whereas the VR-operations are not). The original definitions (see [8]) were given for graphs with multiple edges and hyperedges. In Section 7, we prove that for a set of simple graphs, HR-recognizability is the same in the HR-algebra of simple graphs and in the larger HR-algebra of graphs with multiple edges. Without being extremely difficult, the proof is not just a routine verification.

In Section 8, we consider an algebra arising from the theory of modular decomposition of graphs. We show that under a natural finiteness condition, the corresponding class of recognizable sets is equal to that of VR-recognizable ones.

In an appendix, we clarify the definitions of certain equivalences of logical formulas, focusing on cases where they are decidable, and we give upper bounds to the cardinalities of the quotient sets for these equivalences. These results yield upper bounds to the number of

equivalence classes in logically based congruences. They are thus useful for the investigation of recognizability in view of the cases where the sets under consideration are defined by logical formulas. They also provide elements to appreciate (an upper bound of) the complexity of the algorithms underlying a number of the effective proofs in the main body of the paper.

This work has been presented in invited lectures by Courcelle [17] and Weil [47].

## 2. Recognizability

The notion of a recognizable set is due to Mezei and Wright [37]. It was originally defined for one-sort structures, and we adapt it to many-sorted ones with infinitely many sorts. We begin with definitions concerning many-sorted algebras.

### 2.1. Algebras

We follow essentially the notation and definitions from [45], see also [13]. Let  $S$  be a set called the set of *sorts*. An  $S$ -signature is a set  $\mathcal{F}$  given with two mappings  $\alpha: \mathcal{F} \rightarrow \text{seq}(S)$  (the set of finite sequences of elements of  $S$ ), called the *arity* mapping, and  $\sigma: \mathcal{F} \rightarrow S$ , called the *sort* mapping. We denote by  $\rho(f)$  the length of the sequence  $\alpha(f)$ , which we call also *arity*. The *type* of  $f$  in  $\mathcal{F}$  is the pair  $(\alpha(f), \sigma(f))$  that we shall rather write  $\alpha(f) \rightarrow \sigma(f)$ , or  $(s_1, s_2, \dots, s_n) \rightarrow s$  if  $\alpha(f) = (s_1, \dots, s_n)$  and  $\sigma(f) = s$ . The sequence  $\alpha(f)$  may be empty (that is,  $n = 0$ ), in which case  $f$  is called a *constant of type*  $\sigma(f) = s$ .

An  $\mathcal{F}$ -algebra is an object  $M = \langle (M_s)_{s \in S}, (f_M)_{f \in \mathcal{F}} \rangle$ , where for each  $s \in S$ ,  $M_s$  is a non-empty set, called the *domain of sort*  $s$  of  $M$ . For a nonempty sequence of sorts  $\mu = (s_1, \dots, s_n)$ , we denote by  $M_\mu$  the product  $M_{s_1} \times M_{s_2} \times \dots \times M_{s_n}$ . If  $\rho(f) > 0$ , then  $f_M$  is a total mapping from  $M_{\alpha(f)}$  to  $M_{\sigma(f)}$ . If  $f$  is a constant of type  $s$ , then  $f_M$  is an element of  $M_s$ . The objects  $f_M$  are called the *operations* of  $M$ . We assume that  $M_s \cap M_{s'} = \emptyset$  for  $s \neq s'$ . We also let  $M$  denote the union of the  $M_s$  ( $s \in S$ ). For  $d \in M$ , we let  $\sigma(d)$  denote the unique  $s \in S$  such that  $d \in M_s$ .

A mapping  $h: M \rightarrow M'$  between  $\mathcal{F}$ -algebras is a *homomorphism* (or  $\mathcal{F}$ -homomorphism if it is useful to specify the signature) if it maps  $M_s$  into  $M'_s$  for each sort  $s$  and it commutes with the operations of  $\mathcal{F}$ .

We denote by  $T(\mathcal{F})$  the set of finite well-formed terms built with  $\mathcal{F}$  (we will call them  $\mathcal{F}$ -terms), and by  $T(\mathcal{F})_s$  the set of those terms of sort  $s$  (the sort of a term is that of its leading symbol). If  $\mathcal{F}$  has no constant the set  $T(\mathcal{F})$  is empty.

There is a standard structure of  $\mathcal{F}$ -algebra on  $T(\mathcal{F})$ . Its domain of sort  $s$  is  $T(\mathcal{F})_s$ , and  $T(\mathcal{F})$  can be characterized as the *initial*  $\mathcal{F}$ -algebra. This means that for every  $\mathcal{F}$ -algebra  $M$ , there is a unique homomorphism  $\text{val}_M: T(\mathcal{F}) \rightarrow M$ . If  $t \in T(\mathcal{F})_s$ , the image of  $t$  under  $\text{val}_M$  is an element of  $M_s$ , also denoted by  $t_M$ . It is nothing but the evaluation of  $t$  in  $M$ , where the function symbols are interpreted by the corresponding functions of  $M$ . One can consider  $t$  as a term *denoting*  $t_M$ , and  $t_M$  as the *value* of  $t$  in  $M$ . The set of values in  $M$  of the terms in  $T(\mathcal{F})$  is called the *subset generated* by  $\mathcal{F}$ . We say that a subset of  $M$  is *finitely generated* if it is the set of values of terms in  $T(\mathcal{F}')$  for some finite subset  $\mathcal{F}'$  of  $\mathcal{F}$ .

Let  $\mathcal{F}$  be an  $\mathbf{S}$ -signature,  $\mathcal{F}'$  be an  $\mathbf{S}'$ -signature where  $\mathbf{S}' \subseteq \mathbf{S}$ . We say that  $\mathcal{F}'$  is a *subsignature* of  $\mathcal{F}$ , written  $\mathcal{F}' \subseteq \mathcal{F}$ , if  $\mathcal{F}'$  is a subset of  $\mathcal{F}$  and the types of every  $f$  in  $\mathcal{F}'$  are the same with respect to  $\mathcal{F}$  and to  $\mathcal{F}'$ . We say then that an  $\mathcal{F}'$ -algebra  $M'$  is a *subalgebra* of an  $\mathcal{F}$ -algebra  $M$  if  $M'_s \subseteq M_s$  for every  $s \in \mathbf{S}'$ , and every operation of  $M'$  coincides with the restriction to the domains of  $M'$  of the corresponding operation of  $M$ .

We will often encounter the case where an  $\mathcal{F}$ -algebra  $M$  is also the carrier of a  $\mathcal{G}$ -algebra, and the  $\mathcal{G}$ -operations of  $M$  can be expressed as  $\mathcal{F}$ -terms: in that case, we say that the  $\mathcal{G}$ -operations of  $M$  are  $\mathcal{F}$ -*derived*, and the  $\mathcal{G}$ -algebra  $M$  is an  $\mathcal{F}$ -*derived algebra* (or it is *derived from*  $M$ ).

More formally, an  $\mathbf{S}$ -sorted set of variables is a pair  $(X, \sigma)$  consisting of a set  $X$  and a sort mapping  $\sigma: X \rightarrow \mathbf{S}$  (usually denoted simply by  $X$ ). We let  $T(\mathcal{F}, X)$  be the set of  $(\mathcal{F} \cup X)$ -terms written with  $\mathcal{F} \cup X$ , where it is understood that the variables are among the nullary symbols (constants) of  $\mathcal{F} \cup X$ .  $T(\mathcal{F}, X)_s$  denotes the subset of those terms of sort  $s$ . Now if  $\mathcal{X}$  is a finite sequence of pairwise distinct variables from  $X$  and  $t \in T(\mathcal{F}, X)_s$ , we denote by  $t_{M, \mathcal{X}}$  the mapping from  $M_{\sigma(\mathcal{X})}$  to  $M_s$  associated with  $t$  in the obvious way ( $\sigma(\mathcal{X})$  denotes the sequence of sorts of the elements of  $\mathcal{X}$ ). We call  $t_{M, \mathcal{X}}$  a *derived operation of the algebra*  $M$ . If  $\mathcal{X}$  is known from the context, we write  $t_M$  instead of  $t_{M, \mathcal{X}}$ . This is the case in particular if  $t$  is defined as a member of  $T(\mathcal{F}, \{x_1, \dots, x_k\})$ : the sequence  $\mathcal{X}$  is implicitly  $(x_1, \dots, x_k)$ .

## 2.2. Recognizable subsets

Let  $\mathcal{F}$  be an  $\mathbf{S}$ -signature. An  $\mathcal{F}$ -algebra  $M$  is *locally finite* if each domain  $M_s$  is finite. If  $M$  is an  $\mathcal{F}$ -algebra and  $s \in \mathbf{S}$  is a sort, a subset  $L$  of  $M_s$  is *M-recognizable* if there exists a locally finite  $\mathcal{F}$ -algebra  $A$ , a homomorphism  $h: M \rightarrow A$ , and a (finite) subset  $C$  of  $A_s$  such that  $L = h^{-1}(C)$ .

We denote by  $\text{Rec}(M)_s$  the family of  $M$ -recognizable subsets of  $M_s$ . In some cases it will be useful to stress the relevant signature and we will talk of  $\mathcal{F}$ -recognizable sets instead of  $M$ -recognizable sets.

An equivalent definition can be given in terms of finite congruences. A *congruence* on  $M$  is an equivalence relation  $\approx$  on  $M = \bigcup_{s \in \mathbf{S}} M_s$ , such that each set  $M_s$  is a union of equivalence classes, and which is stable under the operations of  $M$ . It is *locally finite* if for each sort  $s$ , the restriction  $\approx_s$  of  $\approx$  to  $M_s$  has finite index. A congruence *saturates a set* if this set is a union of classes. A subset  $L$  of  $M_s$  is *M-recognizable* if and only if it is saturated by a locally finite congruence on  $M$ .

The following facts are easily verified from the definition of recognizability or its characterization in terms of congruences (see [14]), and will be used freely in the sequel.

**Proposition 2.1.** *Let  $M$  be an  $\mathcal{F}$ -algebra.*

- *For each sort  $s$ , the family  $\text{Rec}(M)_s$  contains  $M_s$  and the empty set, and it is closed under union, intersection and difference.*
- *If  $h$  is a unary derived operation of  $M$  or a homomorphism of  $M'$  into  $M$ , (where  $M'$  is another  $\mathcal{F}$ -algebra), then the inverse image under  $h$  of an  $M$ -recognizable set is recognizable.*

- If  $N$  is a  $\mathcal{G}$ -algebra with the same domain as  $M$ , and if every  $\mathcal{G}$ -congruence of  $N$  is an  $\mathcal{F}$ -congruence of  $M$  (e.g.  $N$  is derived from  $M$ , or  $\mathcal{G}$  is obtained from  $\mathcal{F}$  by adding constants), then every  $M$ -recognizable set is  $N$ -recognizable. If in addition  $\mathcal{G}$  contains  $\mathcal{F}$ , then  $M$  and  $N$  have the same recognizable subsets.
- If  $M'$  is a subalgebra of  $M$  and  $L$  is an  $M$ -recognizable set, then  $L \cap M'$  is  $M'$ -recognizable. This includes the case where  $M'$  has the same domain as  $M$ , and is an  $\mathcal{F}'$ -algebra for some subsignature  $\mathcal{F}'$  of  $\mathcal{F}$ .
- Suppose that  $M$  is generated by  $\mathcal{F}$  and let  $\text{val}_M$  be the evaluation homomorphism from  $T(\mathcal{F})$  onto  $M$ . A subset  $L$  of  $M_s$  is  $\mathcal{F}$ -recognizable if and only if  $\text{val}_M^{-1}(L)$  is a recognizable subset of  $T(\mathcal{F})$ . If in addition  $\mathcal{F}$  is finite, then this is equivalent to the existence of a finite tree-automaton recognizing  $\text{val}_M^{-1}(L)$ .

**Example 2.2.** On the set of all words over a finite alphabet  $A$ , let us consider the binary operation of the concatenation product, and the unary operation  $u \mapsto u^2$ , which is derived from the concatenation product. Then the 3rd statement in Proposition 2.1 shows that we have the same recognizable subsets as if we considered only the concatenation product. It is interesting to note that, in contrast, adding the operation  $u \mapsto u^2$  to the signature adds new equational languages, e.g. the set of all squares.

We will see more technical conditions that guarantee the transfer of recognizability between algebras in Section 2.4 below.

### 2.3. Remarks on the notion of recognizability

We gather here some observations on the significance of recognizability.

First, we note that if  $f$  is an operation of an  $\mathcal{F}$ -algebra  $M$ , with arity  $k$ , and if  $B_1, \dots, B_k$  are  $M$ -recognizable, then  $f(B_1, \dots, B_k)$  is not necessarily recognizable. This is discussed for instance in [10], where sufficient conditions are given to ensure that  $f(B_1, \dots, B_k)$  is recognizable. It is well-known for instance that the product of two recognizable subsets of the free monoid (word languages) or of the trace monoid is recognizable; a similar result holds for recognizable sets of trees.

Now, let  $M$  be an  $\mathcal{F}$ -algebra and let  $\mathcal{F}'$  be a signature which differs from  $\mathcal{F}$  only by the choice of constants and their values. In particular,  $\mathcal{F}'$  may be obtained from  $\mathcal{F}$  by the addition of countably many new constants. Then the congruences on  $M$  are the same with respect to  $\mathcal{F}$  and to  $\mathcal{F}'$  and it follows that a subset of  $M$  is  $\mathcal{F}$ -recognizable if and only if it is  $\mathcal{F}'$ -recognizable.

It is customary to assume that the  $\mathcal{F}$ -algebra  $M$  is generated by the signature  $\mathcal{F}$ . If  $M$  is a countable  $\mathcal{F}$ -algebra that is not generated by  $\mathcal{F}$ , we can enrich  $\mathcal{F}$  to  $\mathcal{F}'$  by adding to  $\mathcal{F}$  one constant of the appropriate sort for each element of  $M$ . Then  $\mathcal{F}'$  generates  $M$  (in a trivial way). As noted above,  $M$  has the same  $\mathcal{F}$ - and  $\mathcal{F}'$ -recognizable subsets. If  $L$  is one of these subsets, the set  $\text{val}_M^{-1}(L)$  of  $\mathcal{F}'$ -terms is recognizable but we cannot do much with it, because we lack the notion of a finite tree-automaton. See the conclusion of the paper for a further discussion of this point.

Finally, we can question the interest of the notion of a recognizable set. Is it interesting in every algebra? The answer is clearly no. Let us explain why.

If the algebraic structure over the considered set  $M$  is poor, for example in the absence of non-nullary functions, then every set  $L$  is recognizable, by a congruence with two classes, namely  $L$  and its complement. The notion of recognizability becomes void.

Another extreme case is when the algebraic structure is so rich that there are very few recognizable sets. For an example, consider the set  $\mathbb{N}$  of natural integers equipped with the successor and the predecessor functions (predecessor is defined by  $\text{pred}(0) = 0$ ,  $\text{pred}(n + 1) = n$ ). The only recognizable sets are  $\mathbb{N}$  and the empty set. Indeed, if  $\sim$  is a congruence and if  $n \sim n + p$  for some  $n \geq 0$ ,  $p > 0$ , then by using the function  $\text{pred}$   $n + p - 1$  times, we find that  $0 \sim 1$ . It follows (using the successor function repeatedly) that any two integers are equivalent.

Intuitively, if one enriches an algebraic structure by adding new operations, one gets fewer recognizable sets.

For another example, let us consider the monoid  $\{a, b\}^*$  of words over two letters. Let us add a unary operation, the *circular shift*, defined by:  $sh(\varepsilon) = \varepsilon$  and  $sh(au) = ua$ ,  $sh(bu) = ub$ , for every word  $u$ . The language  $a^*b$  is no longer recognizable w.r.t. this new structure, however recognizability does not degenerate completely since every commutative language that is recognizable in the usual sense remains recognizable in the enriched algebraic structure.

It is not completely clear yet which algebraic condition makes recognizability “interesting”.

#### 2.4. Technical results on recognizability

The statements in this section explain how to transfer a locally finite congruence from one algebra to another, possibly with a different signature, and hence how to transfer recognizability properties between algebras. Proposition 2.1 above contains examples of such results.

The statements that follow will be used in the proof of some of our main results, in Section 4. They are, unfortunately, heavily technical in their statements (but not in their proofs...).

**Lemma 2.3.** *Let  $\mathcal{F}$  be an  $\mathbf{S}$ -signature and let  $\mathcal{G}$  be a  $\mathbf{T}$ -signature. Let  $S$  be an  $\mathcal{F}$ -algebra and let  $T$  be a  $\mathcal{G}$ -algebra. Let also  $\mathcal{H}$  be a collection  $(\mathcal{H}_{t,s})$  such that, for each  $t \in \mathbf{T}$  and  $s \in \mathbf{S}$ ,  $\mathcal{H}_{t,s}$  consists of mappings from  $T_t$  into  $S_s$  with the following property:*

*for each operation  $g \in \mathcal{G}$  of type  $(t_1, \dots, t_r) \mapsto t$  and for each  $h \in \mathcal{H}_{t,s}$ , there exist sorts  $s_1, \dots, s_r \in \mathbf{S}$ , mappings  $h_i \in \mathcal{H}_{t_i, s_i}$  ( $1 \leq i \leq r$ ) and an  $\mathcal{F}$ -derived operation  $f$  of type  $(s_1, \dots, s_r) \mapsto s$  such that, for every  $x_1 \in T_{t_1}, \dots, x_r \in T_{t_r}$ ,  $h(g(x_1, \dots, x_r)) = f(h_1(x_1), \dots, h_r(x_r))$ .*

*Finally, let  $\equiv$  be an  $\mathcal{F}$ -congruence on  $S$  and let  $\approx$  be the equivalence relation defined, on each  $T_t$ , by*

$$x \approx y \text{ if and only if } h(x) \equiv h(y) \text{ for every } h \in \mathcal{H}_{t,s}, \quad s \in \mathbf{S}.$$

*Then  $\approx$  is a  $\mathcal{G}$ -congruence on  $T$ .*

**Proof.** Let  $g$  be an operation in  $\mathcal{G}$ , of type  $(t_1, \dots, t_r) \mapsto t$ , and let  $x_1, y_1 \in T_{t_1}, \dots, x_r, y_r \in T_{t_r}$  such that  $x_i \approx y_i$  for each  $i = 1, \dots, r$ . Let also  $h \in \mathcal{H}_{t,s}$  with  $s \in S$ .

By hypothesis, there exist sorts  $s_1, \dots, s_r \in S$ , mappings  $h_i \in \mathcal{H}_{t_i, s_i}$  (for  $i = 1, \dots, r$ ) and an  $\mathcal{F}$ -derived operation  $f$  of type  $(s_1, \dots, s_r) \mapsto s$  such that

$$\begin{aligned} h(g(x_1, \dots, x_r)) &= f(h_1(x_1), \dots, h_r(x_r)) \\ h(g(y_1, \dots, y_r)) &= f(h_1(y_1), \dots, h_r(y_r)). \end{aligned}$$

Since  $x_i \approx y_i$  for each  $i$ , we have  $h_i(x_i) \equiv h_i(y_i)$ ; and since  $\equiv$  is an  $\mathcal{F}$ -congruence, it follows that  $h(g(x_1, \dots, x_r)) \equiv h(g(y_1, \dots, y_r))$ . Thus we have  $g(x_1, \dots, x_r) \approx g(y_1, \dots, y_r)$ , which concludes the proof.  $\square$

With the notation of Lemma 2.3, for each sort  $t \in T$ , let  $\leq_t$  be the quasi-order relation defined on  $\mathcal{H}_t = \bigcup_{s \in S} \mathcal{H}_{t,s}$  by

$$h \leq_t h' \text{ if there exists an } \mathcal{F}\text{-derived unary operation } f \text{ such that } h' = f \circ h.$$

**Lemma 2.4.** *With the notation of Lemma 2.3, if for each  $t$  the order relation associated with  $\leq_t$  has a finite number of minimal elements, and if the  $\mathcal{F}$ -congruence  $\equiv$  on  $S$  is locally finite, then the  $\mathcal{G}$ -congruence  $\approx$  on  $T$  is locally finite.*

**Proof.** Let  $t \in T$ . We want to show that there are only finitely many  $\approx$ -classes in  $T_t$ . By assumption, there exist elements  $h_1, \dots, h_k \in \mathcal{H}_t$  such that every mapping of  $\mathcal{H}_t$  is of the form  $f \circ h_i$  for some  $1 \leq i \leq k$  and some  $\mathcal{F}$ -derived operation  $f$ .

For each  $i$ , let  $S_{s_i}$  be the range of  $h_i$  and let  $n_i$  be the number of  $\equiv$ -classes in  $S_{s_i}$ . It is immediately verified from the definition of  $\leq_t$  that if  $x, y \in T_t$ , then  $x \approx y$  if and only if  $h_i(x) \equiv h_i(y)$  for each  $1 \leq i \leq k$ . In particular,  $T_t$  has at most  $n_1 \cdots n_k$   $\approx$ -classes, which concludes the proof.  $\square$

We will actually need even more technical versions of these lemmas.

**Lemma 2.5.** *Let  $S, T, \mathcal{F}, \mathcal{G}$  and  $\mathcal{H}$  be as in Lemma 2.3, and let  $\zeta$  be a  $\mathcal{G}$ -congruence on  $T$  such that:*

*for each operation  $g \in \mathcal{G}$  of type  $(t_1, \dots, t_r) \mapsto t$ , for each  $h \in \mathcal{H}_{t,s}$  and for each  $\vec{z} = (z_1, \dots, z_r)$  where each  $z_i$  is a  $\zeta$ -class of  $T_{t_i}$ , there exist sorts  $s_{1,\vec{z}}, \dots, s_{r,\vec{z}} \in S$ , mappings  $h_{i,\vec{z}} \in \mathcal{H}_{t_i, s_{i,\vec{z}}}$  ( $1 \leq i \leq r$ ) and an  $\mathcal{F}$ -derived operation  $f_{\vec{z}}$  of type  $(s_{1,\vec{z}}, \dots, s_{r,\vec{z}}) \mapsto s$  such that, in  $T$ ,  $h(g(x_1, \dots, x_r)) = f_{\vec{z}}(h_{1,\vec{z}}(x_1), \dots, h_{r,\vec{z}}(x_r))$  if each  $x_i$  is in  $z_i$ . Finally, let  $\equiv$  be an  $\mathcal{F}$ -congruence on  $S$  and let  $\approx$  be the equivalence relation defined, on each  $T_t$ , by*

$$x \approx y \text{ if and only if } x \zeta y \text{ and } h(x) \equiv h(y) \text{ for every } h \in \mathcal{H}_{t,s}, \quad s \in S.$$

*Then  $\approx$  is a  $\mathcal{G}$ -congruence on  $T$ . Moreover, if  $\mathcal{H}$  satisfies the hypothesis of Lemma 2.4 and  $\equiv$  and  $\zeta$  are locally finite, then  $\approx$  is locally finite as well.*

**Proof.** The proof is the same as for Lemmas 2.3 and 2.4.  $\square$

### 3. Algebras of relational structures

Even though we are ultimately interested in studying sets of graphs, it will be convenient to handle the more general case of relational structures. Furthermore, relational structures can be identified with simple directed hypergraphs. Such hypergraphs form a natural representation of terms. See for instance the chapter on hypergraphs in [15] for applications.

In this paper, all graphs and structures are finite or countable. Our proofs will not usually depend on cardinality assumptions on the graphs or structures, and hence our results will hold for finite as well as for infinite graphs or structures. However, recognizability in the algebraic sense we defined, is really interesting only for dealing with finitely generated objects, and hence for finite graphs and structures. For dealing with infinite words, trees and graphs, other tools are necessary, see for instance [40,43,29,30].

#### 3.1. Relational structures

Let  $R$  be a finite set of relation symbols, and  $C$  be a finite set of nullary symbols. Each symbol  $r \in R$  has an associated positive integer called its *rank*, denoted by  $\rho(r)$ . An  $(R, C)$ -structure is a tuple  $S = \langle D_S, (r_S)_{r \in R}, (c_S)_{c \in C} \rangle$  such that  $D_S$  is a (possibly empty) set called the *domain* of  $S$ , each  $r_S$  is a  $\rho(r)$ -ary relation on  $D_S$ , i.e., a subset of  $D_S^{\rho(r)}$ , and each  $c_S$  is an element of  $D_S$ , called the *c-source* of  $S$ .

We denote by  $StS(R, C)$  the class of (finite or countable)  $(R, C)$ -structures, and we sometimes write  $StS(R)$  for  $StS(R, \emptyset)$ . By convention, isomorphic structures will be considered as equal. In the notation  $StS$ ,  $St$  stands for *structures*, while the second  $S$  stands for *sources*.

A structure  $S \in StS(R, C)$  is *source-separated* if  $c_S \neq c'_S$  for  $c \neq c'$ . We will denote by  $StS_{sep}(R, C)$  the class of source-separated structures in  $StS(R, C)$ . See Corollary 3.11 and Section 3.5.2.

In order to handle graphs, we will consider particular kinds of structures in the sequel. We let  $E = \{\text{edge}\}$  be the set of relation symbols consisting of a single binary relation *edge*, intended to represent directed edges. Thus graphs can be seen as the elements of  $StS(E)$ , also written *Graph*. Clearly these graphs are directed, simple (we cannot represent multiple edges) and they may have loops. For a discussion of graphs with multiple edges, see Section 7.

We let  $\mathcal{GS}(C)$  denote the set  $StS(E, C)$ . These structures are called *graphs with sources*. We let  $\mathcal{GS}_{sep}(C)$  denote the intersection  $\mathcal{GS}(C) \cap StS_{sep}(R, C)$ .

We will discuss also *graphs with ports* (Section 4): if  $P$  is a finite set of unary relation symbols called *port labels*, then we denote by  $E_P$  the set of relational symbols  $E \cup P$  and by  $\mathcal{GP}(P)$  the class  $StS(E_P)$ . Port labels are useful for studying the clique-width of graphs, see [18,19] and Remark 4.11.

#### 3.2. The algebra $StS$

We first define some operations on structures.

*Disjoint union.* Let  $C$  and  $C'$  be disjoint sets of constants and let  $S \in \text{StS}(R, C)$  and  $S' \in \text{StS}(R', C')$ . Let us also assume that  $S$  and  $S'$  have disjoint domains. We denote by  $S \oplus S'$  the union of  $S$  and  $S'$ , which is naturally a structure in  $\text{StS}(R \cup R', C \cup C')$ .

If  $S$  and  $S'$  are not disjoint, we replace  $S'$  by a disjoint copy. We need not be very precise on how to choose this copy because different choices will yield isomorphic  $\oplus$ -sums, and we are interested in structures up to isomorphism.

**Remark 3.1.** It is also possible to define a similar operation, without the restriction that  $C$  and  $C'$  are disjoint (as in, say, [9,14]). See Section 3.5.1 for a discussion.

*Quantifier-free definable operations.* Our purpose is now to define functions from  $\text{StS}(R, C)$  to  $\text{StS}(R', C')$  by quantifier-free formulas. We denote by  $QF(R, C, \{x_1, \dots, x_n\})$  the set of quantifier-free formulas on  $(R, C)$ -structures with variables in  $\{x_1, \dots, x_n\}$ .

A *qfd operation scheme* from  $\text{StS}(R, C)$  to  $\text{StS}(R', C')$  is a tuple

$$(\delta, (\varphi_r)_{r \in R'}, (\kappa_{c,d})_{c \in C, d \in C'}),$$

where  $\delta \in QF(R, C, \{x\})$ ,  $\varphi_r \in QF(R, C, \{x_1, \dots, x_{\rho(r)}\})$  if  $r$  is a  $\rho(r)$ -ary relation symbol,  $\kappa_{c,d} \in QF(R, C, \emptyset)$ , such that the following formulas are valid in every structure in  $\text{StS}(R, C)$ , for all  $c, c' \in C$ ,  $d \in C'$  and  $r \in R'$  of arity  $\rho(r)$ :

- $\kappa_{c,d} \wedge \kappa_{c',d} \implies c = c'$ ;
- $\bigvee_{e \in C} \kappa_{e,d}$ ;
- $\kappa_{c,d} \implies \delta(c)$ ;
- $\forall x_1, \dots, x_{\rho(r)} \left( \varphi_r(x_1, \dots, x_{\rho(r)}) \implies \bigwedge_{i=1}^{\rho(r)} \delta(x_i) \right)$ .

The reason for these conditions becomes apparent with the following definition of the *qfd operation*  $g: \text{StS}(R, C) \rightarrow \text{StS}(R', C')$  defined by such a scheme. Let  $S \in \text{StS}(R, C)$ . The domain of the structure  $g(S)$  is the subset of the domain of  $S$  defined by formula  $\delta$  and the relation  $r$  ( $r \in R'$ ) on  $g(S)$  is described by formula  $\varphi_r$ . Finally, if  $d \in C'$ , then  $d_{g(S)} = c_S$  if  $c \in C$  and  $S$  satisfies  $\kappa_{c,d}$ . The first two conditions imposed above assert that relative to  $S$ ,  $c$  is uniquely defined for each  $d$ , the third condition asserts that  $d_{g(S)}$  always lies in the domain of  $g(S)$ , and the fourth condition asserts that the relation  $\varphi_r$  ( $r \in R'$ ) can only relate elements of the domain of  $g(S)$ .

**Remark 3.2.** Note that in the first condition,  $c = c'$  does not mean that  $c$  and  $c'$  are the same constant, but that they have the same value in the considered structure.

**Remark 3.3.** The conditions to be verified by a qfd operation scheme are decidable. It follows that the notion of a qfd operation scheme is effective. See Appendix A (Remark A.4 in particular) for a discussion of this decidability result.

**Example 3.4.** Let  $R$  be a finite set of relational symbols,  $C$  be a finite set of source labels and let  $a, b$  be source labels. We define the following operations.

- if  $a \in C$  and  $b \notin C$ ,  $\text{srcren}_{a \rightarrow b}$  is the unary operation of type  $(R, C) \rightarrow (R, C \setminus \{a\} \cup \{b\})$  which renames the  $a$ -source of a structure to a  $b$ -source;
- if  $a \in C$ ,  $\text{srcfg}_a$  is the unary operation of type  $(R, C) \rightarrow (R, C \setminus \{a\})$  which forgets the  $a$ -source of a structure;

- if  $a \neq b \in C$ ,  $\text{fus}_{a,b}$  is the unary operation of type  $(R, C) \rightarrow (R, C)$  which identifies the  $a$ -source and the  $b$ -source of a structure (so the resulting domain element is both the  $a$ -source and the  $b$ -source), and reorganizes the tuples of the relational structure accordingly.

Note that the operation names  $\text{srcren}_{a \rightarrow b}$ ,  $\text{srcfg}_a$  and  $\text{fus}_{a,b}$  are overloaded: they denote different operations when the sets  $R$  and  $C$  are allowed to vary. A completely formal definition would use operation names such as  $\text{srcren}_{a \rightarrow b, R, C}$ , which would be inconvenient.

It is immediately verified that the operations of the form  $\text{srcren}_{a \rightarrow b}$  and  $\text{srcfg}_a$  are qfd. It is probably worth showing explicitly a qfd operation scheme defining the operation  $\text{fus}_{a,b}$ .

Let  $\delta(x)$  be the formula  $(a = b) \vee ((a \neq b) \wedge (x \neq a))$ . If  $r \in R$  has arity  $\rho(r) = n$ , let  $\varphi_r(x_1, \dots, x_n)$  be the formula

$$\left( (a = b) \wedge r(x_1, \dots, x_n) \right) \vee \left( (a \neq b) \wedge \bigvee_{I \subseteq \{1, \dots, n\}} \left( \bigwedge_{i \in I} (x_i = b) \wedge \bigwedge_{i \notin I} (x_i \neq b) \wedge r(y_1, \dots, y_n) \right) \right),$$

where for each  $I$ ,  $y_i = a$  if  $i \in I$  and  $y_i = x_i$  otherwise. For each  $d \in C$  such that  $d \neq a$  and for each  $c \in C$ , let  $\kappa_{c,d}$  be the formula  $c = d$ ; let  $\kappa_{b,a}$  be the formula  $\text{true}$ , and let  $\kappa_{c,a}$  be the formula  $c = a$  for each  $c \neq b$ . It is now routine to verify that the scheme  $(\delta, (\varphi_r)_{r \in R}, (\kappa_{c,d})_{c,d \in C})$  defines  $\text{fus}_{a,b}$ .

**Remark 3.5.** There is no qfd operation from  $\text{StS}(R)$  into  $\text{StS}(R', C)'$  if  $C' \neq \emptyset$ , because in the absence of constants in the input structure, we cannot define constants in the output structure.

**Example 3.6.** The natural inclusion of  $\text{StS}(R, C)$  into  $\text{StS}(R', C)$  when  $R'$  contains  $R$  is a qfd operation in natural way: the formulas intended to define relations in  $R' \setminus R$  are taken to be identically false.

*The signature  $\mathcal{S}$ .* We define the algebra  $\text{StS}$  of structures with sources as follows. First, let us fix once and for all a countable set of relation symbols containing `edge` and countably many relation symbols of each arity, and a countable set of constants. In the sequel, finite sets of relation symbols  $R$  and finite sets of constants  $C$  will be taken in these fixed sets. The set of sorts consists of all such pairs  $(R, C)$ . The set of elements of  $\text{StS}$  of sort  $(R, C)$  is  $\text{StS}(R, C)$ .

The signature  $\mathcal{S}$  consists of the following operations (interpreted in  $\text{StS}$ ). First, for each pair of sorts  $(R, C)$  and  $(R', C')$  such that  $C \cap C' = \emptyset$ , the disjoint union  $\oplus$  is an operation of type  $((R, C), (R', C')) \rightarrow (R \cup R', C \cup C')$ . Note that we overload the symbol  $\oplus$ , that is, we denote in the same way an infinite number of operations on  $\text{StS}$ . Next, every qfd operation is a (unary) operation in  $\mathcal{S}$ .

Finally, we observe that the signature  $\mathcal{S}$  contains the natural inclusions of  $\text{StS}(R, C)$  into  $\text{StS}(R', C)$  when  $R'$  contains  $R$ , which are qfd (Example 3.6).

As for constants in  $\mathcal{S}$ , one can pick a single source label  $a$ , and consider a single constant  $a$ , denoting the structure with a single element, which is an  $a$ -source, and no relations. Together with the operations in  $\mathcal{S}$ , this constant suffices to generate all finite relational

structures. As noted in Section 2.3, the choice of constants does not affect recognizability. It only affects the generating power of the signature, but this is not our point in this paper.

### 3.3. Elementary properties of $StS$

We first consider the composition of qfd operations.

**Proposition 3.7.** *Qfd operations in  $StS$  are closed under composition (whenever types fit for defining meaningful composition).*

**Proof.** Let  $g: StS(R, C) \rightarrow StS(R', C')$  and  $g': StS(R', C') \rightarrow StS(R'', C'')$  be qfd operations, given respectively by the schemes  $(\delta, (\psi_r)_{r \in R'}, (\kappa_{c,d})_{c \in C', d \in C'})$  and  $(\delta', (\psi'_r)_{r \in R''}, (\kappa'_{c,d})_{c \in C'', d \in C''})$ .

The composition  $g' \circ g$  turns an  $(R, C)$ -structure into an  $(R'', C'')$ -structure.

Let  $\delta^0, \psi_r^0$  ( $r \in R''$ ) and  $\kappa_{c,d}^0$  ( $c \in C', d \in C''$ ) be obtained from  $\delta', \psi'_r$  and  $\kappa'_{c,d}$  by replacing every occurrence of  $r(y_1, \dots, y_{\rho(r)})$  ( $r \in R'$ ) by  $\psi_r(y_1, \dots, y_{\rho(r)})$ ; our formulas are now in the language of  $(R, C')$ -structures and we need to “translate” the constants  $d \in C'$  into elements of  $C$ . However, this translation, a mapping from  $C'$  to  $C$ , depends on the structure in which we operate.

To reflect this observation, for each mapping  $h: C' \rightarrow C$ , we let  $h(\delta^0)$  be the conjunction of the formulas  $\kappa_{h(d),d}$  ( $d \in C'$ ) and the formula obtained from  $\delta^0$  by replacing each occurrence of  $d$  ( $d \in C'$ ) by  $h(d)$ . Finally, we let  $\delta''$  be the disjunction of the  $h(\delta^0)$  when  $h$  runs over all mappings from  $C'$  to  $C$ .

We proceed in the same fashion to define  $\psi''_r$  and  $\kappa''_{c,d}$  for each  $r \in R''$  and each  $c \in C', d \in C''$ . Finally, if  $b \in C$  and  $d \in C''$ , we let  $\lambda_{b,d} = \bigvee_{c \in C'} (\kappa'_{b,c} \wedge \kappa''_{c,d})$ .

It is a routine verification that  $(\delta'', (\psi''_r)_{r \in R''}, (\lambda_{b,d})_{b \in C, d \in C''})$  is a qfd operation scheme, which defines the operation  $g' \circ g$ . This completes the proof.  $\square$

For each  $S \in StS(R, C)$ , we define the *type* of  $S$ , written  $\zeta(S)$ , to be the restriction of  $S$  to its set of sources. That is: the domain of  $\zeta(S)$  is the set of  $C$ -sources of  $S$ , and the relations of  $\zeta(S)$  are those tuples of  $C$ -sources that are relations in  $S$ . In order to simplify notation, we also denote by  $\zeta$  the equivalence relation on  $StS$  given by

$$S \zeta T \quad \text{if and only if} \quad \zeta(S) \text{ and } \zeta(T) \text{ are isomorphic.}$$

**Lemma 3.8.** *Let  $S, T \in StS(R, C)$ . Then  $S \zeta T$  if and only if  $S$  and  $T$  satisfy the same formulas in  $QF(R, C, \emptyset)$ .*

**Proof.** A formula in  $QF(R, C, \emptyset)$  is a Boolean combination of atoms of the form  $c = d$  where  $c, d \in C$ , or  $r(x_1, \dots, x_n)$  where  $r \in R$  has arity  $n$  and the  $x_i$  are in  $C$ . It is immediate that such an atom is true in  $S$  if and only if it is true in  $\zeta(S)$ . Thus  $S$  and  $\zeta(S)$  satisfy the same formulas in  $QF(R, C, \emptyset)$ : in particular,  $\zeta$ -equivalent structures satisfy the same formulas in  $QF(R, C, \emptyset)$ . Thus, if we denote by  $Th_{0,R,C}^{FO}(S)$  the set of formulas in  $QF(R, C, \emptyset)$  that are satisfied by  $S$  (see Section 3.4), we find that  $Th_{0,R,C}^{FO}(S) = Th_{0,R,C}^{FO}(\zeta(S))$ .

Conversely, we observe that if  $S$  is a structure in  $\mathcal{StS}(R, C)$ , which consists only of its  $C$ -sources (that is,  $S = \zeta(S)$ ), then  $S$  is entirely described by some formula in  $QF(R, C, \emptyset)$ . Thus, if  $\zeta(S) \neq \zeta(T)$ , then  $Th_{0,R,C}^{FO}(S) \neq Th_{0,R,C}^{FO}(T)$ . This suffices to conclude the proof.  $\square$

The type relation  $\zeta$  has the following important property.

**Proposition 3.9.** *The type relation  $\zeta$  is a locally finite congruence on  $\mathcal{StS}$ .*

**Proof.** The verification that  $\zeta(S \oplus S') = \zeta(S) \oplus \zeta(S')$  ( $S \in \mathcal{StS}(R, C)$ ,  $S' \in \mathcal{StS}(R', C')$  and  $C \cap C' = \emptyset$ ) is immediate. Let us now consider a qfd operation  $g: \mathcal{StS}(R, C) \rightarrow \mathcal{StS}(R', C')$ , specified by the qfd operation scheme  $(\delta, (\psi_r)_{r \in R'}, (\kappa_{c,d})_{c \in C, d \in C'})$ . By Lemma 3.8,  $S$  and  $\zeta(S)$  satisfy the same formulas of  $QF(R, C, \emptyset)$ . In particular, for each  $c \in C$  and  $d \in C'$ ,  $S$  and  $\zeta(S)$  both satisfy  $\kappa_{c,d}$ , or both satisfy its negation. Thus  $g(S)$  and  $g(\zeta(S))$  have the same sources, and hence  $\zeta(g(S)) = \zeta(g(\zeta(S)))$ .

We have just shown that the type relation is a congruence. To complete the proof, it suffices to show that for each sort  $(R, C)$ , the set of types of sort  $(R, C)$ , that is, the set  $\zeta(\mathcal{StS}(R, C))$  is finite. Note that if  $S \in \mathcal{StS}(R, C)$ , then  $\zeta(S)$  has cardinality at most  $\text{card}(C)$  (and also at most  $\text{card}(S)$ ). It follows that  $\text{card}(\zeta(\mathcal{StS}(R, C))) \leq \text{card}(C)! \prod_{r \in R} 2^{\text{card}(C)^{\rho(r)}}$ .  $\square$

**Remark 3.10.** Proposition 3.9 can be seen as a particular case of a result of Feferman and Vaught [25], Theorem 3.12 below, which will be used in Section 6. The simple formulation above will be very useful.

Note that the knowledge of  $\zeta(S)$  is sufficient to determine whether  $S$  is a source-separated structure. This observation is used to prove the following corollary.

**Corollary 3.11.** *Let  $(R, C)$  be a sort in  $\mathcal{StS}$ . Then  $\mathcal{StS}_{\text{sep}}(R, C)$  is a recognizable subset of  $\mathcal{StS}(R, C)$ .*

**Proof.** Whether a structure  $S$  is source-separated depends only on its type  $\zeta(S)$ : in particular, the type congruence  $\zeta$  saturates  $\mathcal{StS}_{\text{sep}}(R, C)$ . By Proposition 3.9, this relation is a locally finite congruence, and hence  $\mathcal{StS}_{\text{sep}}(R, C)$  is recognizable.  $\square$

### 3.4. A result of Feferman and Vaught

If  $(R, C)$  is a sort of  $\mathcal{StS}$ , we denote by  $FO(R, C)$  the set of closed first-order formulas over  $R$  and  $C$ . For each integer  $d$ , we denote by  $FO_d(R, C)$  the set of those formulas of quantifier-depth at most  $d$ . Up to a decidable syntactic equivalence (taking into account Boolean laws, properties of equality, renaming of quantified variables, see Appendix A), there are only finitely many formulas in each set  $FO_d(R, C)$ . Thus, we can reason as if  $FO_d(R, C)$  was actually finite.

For an  $(R, C)$ -structure  $S$ , we let its  $FO_d$ -theory be the set  $Th_{d,R,C}^{FO}(S)$  of formulas in  $FO_d(R, C)$  that are valid in  $S$ . It is finite since it is a subset of the finite set  $FO_d(R, C)$ .

**Theorem 3.12.** *Let  $d \geq 0$ .*

- (1) *For every qfd operation  $f$  of type  $(R, C) \rightarrow (R', C')$ , there exists a mapping  $f_d^\#$  such that, for every  $(R, C)$ -structure  $S$*

$$Th_{d,R',C'}^{FO}(f(S)) = f_d^\#(Th_{d,R,C}^{FO}(S)).$$

- (2) *For every  $(R, C)$  and  $(R', C')$ , where  $C$  and  $C'$  are disjoint, there exists a binary function  $\oplus_d^\#$  such that, for every  $(R, C)$ -structure  $S$ , and every  $(R', C')$ -structure  $S'$ ,*

$$Th_{d,R \cup R', C \cup C'}^{FO}(S \oplus S') = Th_{d,R,C}^{FO}(S) \oplus_d^\# Th_{d,R',C'}^{FO}(S').$$

**Remark 3.13.** The second assertion was proved in [25] for first-order logic, and extended by Shelah to monadic second-order logic [42]. The importance of this result is discussed by Makowsky in [34].

**Remark 3.14.** The functions  $f_d^\#$  and  $\oplus_d^\#$  have finite domains and codomains. However these sets are quite large. These functions can be (at least in principle) effectively determined for given  $(R, C)$ ,  $(R', C')$ , and  $d$ .

### 3.5. Variants of the algebra of relational structures

In the literature on recognizable and equational sets of graphs, several variants of the signature  $S$  and the algebra  $StS$  are considered, notably a variant where the definition of the disjoint union is replaced by a more general parallel product, and a variant where all structures are assumed to be source-separated. We verify in this section that these variants do not yield different notions of recognizability.

#### 3.5.1. Parallel composition vs. disjoint union

In the literature (e.g. [9,14]), the operation of disjoint union  $\oplus$  is sometimes replaced by the so-called *parallel composition* (or *product*), written  $\parallel$ , an operation of type  $((R, C), (R', C')) \rightarrow (R \cup R', C \cup C')$  for which we do not assume that  $C$  and  $C'$  are disjoint. If  $S \in StS(R, C)$  and  $S' \in StS(R', C')$ , the parallel composition  $S \parallel S'$  is obtained by taking the (set-theoretic) disjoint union of  $S$  and  $S'$  and then identifying the  $c$ -sources of  $S$  and  $S'$  for each  $c \in C \cap C'$ . Let  $S_\parallel$  denote the signature obtained from  $S$  by substituting  $\parallel$  for  $\oplus$ .

**Proposition 3.15.** *Let  $L$  be a subset of  $StS$ . Then  $L$  is  $S$ -recognizable if and only if it is  $S_\parallel$ -recognizable.*

**Proof.** We first observe that the operation  $\oplus$  is a particular case of  $\parallel$ . Therefore  $S$  is a sub-signature of  $S_\parallel$  and hence, every  $S_\parallel$ -recognizable set is  $S$ -recognizable.

To prove the converse, it suffices to verify that  $\parallel$  is an  $S$ -derived operation by Proposition 2.1. Indeed, if  $S \in StS(R, C)$  and  $S' \in StS(R', C')$ , the parallel composition  $S \parallel S'$  can be obtained by the following sequence of  $S$ -operations (see Example 3.4 for their definition):

- for each  $c \in C \cap C'$ , apply the qfd operation  $src\text{ren}_{c \rightarrow \bar{c}}$  which renames the  $c$ -source in  $S'$  with a new source label, say  $\bar{c}$ , not in  $C$ ; let  $\bar{S}'$  be the resulting structure;

- take the disjoint union  $S \oplus \bar{S}'$ ;
- for each  $c \in C \cap C'$ , apply the operation  $\text{fus}_{c,\bar{c}}$  which identifies the  $c$ -source and the  $\bar{c}$ -source in  $S \oplus \bar{S}'$ ;
- apply the source-forgetting operation  $\text{srcfg}_{\bar{c}}$  for each  $c \in C \cap C'$ .  $\square$

### 3.5.2. Source-separated structures

The property that  $c_S \neq c'_S$  for  $c \neq c'$  is called *source separation*. This property makes it easier to work with operations on structures and graphs, and hence we discuss a variant of the  $\mathcal{S}$ -algebra  $\mathcal{StS}$ , which handles source-separated structures. We will also use it in Section 6.

Recall that  $\mathcal{StS}_{\text{sep}}(R, C)$  denotes the set of source-separated structures in  $\mathcal{StS}(R, C)$ . We now define a subsignature  $\mathcal{S}_{\text{sep}}$  of  $\mathcal{S}$  such that  $\mathcal{StS}_{\text{sep}}$  is a sub-algebra of  $\mathcal{StS}$ .

Disjoint union  $\oplus$  clearly preserves source separation, and is part of  $\mathcal{S}_{\text{sep}}$ . Next we include in  $\mathcal{S}_{\text{sep}}$  the operations specified by qfd operation schemes such that, for each  $c \in C$  and  $d \neq d' \in C'$  (see the notation in Section 3.2),

$$\kappa_{c,d} \implies \neg \kappa_{c,d'}, \quad (1)$$

which guarantees that the operation preserves source separation.

**Example 3.16.** The operations  $\text{srcren}_{a \rightarrow b}$  and  $\text{srcfg}_a$  defined in Example 3.4 are in  $\mathcal{S}_{\text{sep}}$ . The operation  $\text{fus}_{a,b}$  defined in the same example is not.

In contrast, the operation written  $\text{fus}_{a \rightarrow b}$ , which identifies the  $a$ -source and the  $b$ -source of a structure as in  $\text{fus}_{a,b}$ , and makes the resulting element of the domain a  $b$ -source but not an  $a$ -source, preserves source separation. It can be written as  $\text{fus}_{a \rightarrow b} = \text{srcfg}_a \circ \text{fus}_{a,b}$ .

The operation which, given a graph with source labels  $a$  and  $b$ , exchanges the source labels  $a$  and  $b$  if the corresponding vertices are linked by an edge and does nothing otherwise, is another example of a qfd operation in  $\mathcal{S}_{\text{sep}}$ .

Regarding the effectiveness of the definition of  $\mathcal{S}_{\text{sep}}$ , we observe the following.

**Proposition 3.17.** *Given a qfd operation scheme, one can decide whether the corresponding qfd operation preserves source separation.*

**Proof.** Let  $g$  be the qfd operation specified by the given qfd operation scheme, and let  $\mathcal{StS}(R, C)$  be the domain of  $g$ . One can effectively construct the images under  $g$  of every type in  $\mathcal{StS}(R, C)$ , since there are only finitely many of them, and they can all be enumerated. One can then verify whether the operation preserves source-separation on types.

Now it follows from the proof of Proposition 3.9 that for each  $S \in \mathcal{StS}(R, C)$ , we have  $\zeta(g(\zeta(S))) = \zeta(g(S))$ . In particular,  $g$  preserves source separation if and only if it preserves it for the structures of the form  $\zeta(S)$ . Thus one can effectively decide whether  $g \in \mathcal{S}_{\text{sep}}$ .  $\square$

We now show that the restriction to source-separated structures does not change the notion of recognizability.

**Theorem 3.18.** *Let  $L$  be a subset of  $St\mathcal{S}_{\text{sep}}$ . Then  $L$  is  $\mathcal{S}$ -recognizable if and only if it is  $\mathcal{S}_{\text{sep}}$ -recognizable.*

**Proof.** By definition,  $\mathcal{S}_{\text{sep}}$  is a subsignature of  $\mathcal{S}$ , so every  $\mathcal{S}$ -recognizable set is  $\mathcal{S}_{\text{sep}}$ -recognizable.

To prove the converse, we first define a mapping  $h$ , which maps a structure  $S \in St\mathcal{S}(R, C)$  to a source-separated structure  $h(S) \in St\mathcal{S}_{\text{sep}}(R, C)$  by splitting sources that were identified in  $S$ .

We assume that the countable set of constant symbols (from which  $C$  is taken, see Section 3.2) is linearly ordered. Let  $h_0^S: C \rightarrow C$  be given by

$$h_0^S(c) = \min\{d \in C \mid c_S = d_S\}.$$

We let  $C_0^S = h_0^S(C)$  and  $C_1^S = C \setminus C_0^S$ . The structure  $h(S)$  has domain set the disjoint union of  $S$  and  $C_1^S$ . For each  $c \in C_0^S$ , the  $c$ -source of  $h(S)$  is the  $c$ -source of  $S$ , and for each  $c \in C_1^S$ , the  $c$ -source of  $h(S)$  is the element  $c \in C_1^S$ . Finally, for each  $r \in R$ , the relation  $r_{h(S)}$  equals the relation  $r_S$  (so it does not involve the elements of  $C_1^S$ ). Observe that  $h$  is not a qfd operation, and that  $h_0^S$ ,  $C_0^S$  and  $C_1^S$  depend only on  $\zeta(S)$ .

Now let  $L$  be an  $\mathcal{S}_{\text{sep}}$ -recognizable subset of  $St\mathcal{S}_{\text{sep}}$  and let  $\equiv$  be a locally finite  $\mathcal{S}_{\text{sep}}$ -congruence recognizing it. We need to construct a locally finite  $\mathcal{S}$ -congruence  $\sim$  on  $St\mathcal{S}$  which recognizes  $L$ .

The relation  $\sim$  on  $St\mathcal{S}$  is defined as follows. If  $S, T \in St\mathcal{S}(R, C)$ , we say that  $S \sim T$  if  $\zeta(S) = \zeta(T)$  and  $h(S) \equiv h(T)$ . It is immediately verified that  $\sim$  is an equivalence relation. Moreover, the  $\sim$ -class of a structure  $S$  is determined by its  $\zeta$ -class, and by the  $\equiv$ -class of  $h(S)$ . Since both  $\zeta$  and  $\equiv$  are locally finite,  $\sim$  also is locally finite.

Let us now prove that  $\sim$  is an  $\mathcal{S}$ -congruence. Let  $S \sim T \in St\mathcal{S}(R, C)$  and  $S' \sim T' \in St\mathcal{S}(R', C')$ , with  $C \cap C' = \emptyset$ . By Proposition 3.9,  $\zeta(S \oplus S') = \zeta(T \oplus T')$ . It is not difficult to verify that

$$h(S \oplus S') = h(S) \oplus h(S').$$

It follows that  $h(S \oplus S') \equiv h(T \oplus T')$  since  $\oplus$  is an operation in  $\mathcal{S}_{\text{sep}}$ . Thus  $S \oplus S' \sim T \oplus T'$ .

Next let  $g$  be a qfd operation from  $St\mathcal{S}(R, C)$  to  $St\mathcal{S}(Q, B)$ , given by the qfd operation scheme  $(\delta, (\psi_q)_{q \in Q}, (\kappa_{c,b})_{c \in C, b \in B})$ . Let  $S$  and  $T$  be  $\sim$ -equivalent elements of  $St\mathcal{S}(R, C)$ , which will remain fixed for the rest of this proof. We need to show that  $g(S) \sim g(T)$ . We already know from Proposition 3.9 that if  $S \sim T \in St\mathcal{S}(R, C)$ , then  $\zeta(g(S)) = \zeta(g(T))$ , and we want to show that  $h(g(S)) \equiv h(g(T))$ .

Since  $\zeta(g(S)) = \zeta(g(T))$ , the mappings  $h_0^{g(S)}$  and  $h_0^{g(T)}$ , from  $B$  to  $B$ , coincide. Let  $B_0 = h_0^{g(S)}(B)$  and  $B_1 = B \setminus B_0$ . Without loss of generality, we may assume that  $B_1 \cap C = \emptyset$ . The domain set of  $h(g(S))$  (resp.  $h(g(T))$ ) is the disjoint union of the domain of  $g(S)$  (resp.  $g(T)$ ) and  $B_1$ .

It suffices to show that there exists a qfd operation  $k \in \mathcal{S}_{\text{sep}}$ , depending on  $g$  and  $\zeta(S)$ , such that  $h(g(S)) = k(h(S) \oplus B_1)$  and  $h(g(T)) = k(h(T) \oplus B_1)$  (where  $B_1$  is the source-only element of  $St\mathcal{S}_{\text{sep}}(\emptyset, B_1)$ ). Indeed, the fact that  $\equiv$  is an  $\mathcal{S}_{\text{sep}}$ -congruence will then imply that  $h(g(S)) \equiv h(g(T))$ .

Let  $\delta'$  be obtained from  $\delta$  by replacing every occurrence of  $c \in C$  by  $h_0^S(c)$ . For each  $q \in Q$ ,  $c \in C$  and  $b \in B$ , let  $\psi'_q$  be obtained from  $\psi_q$  and  $\kappa'_{c,b}$  be obtained from  $\kappa_{c,b}$  in the same fashion.

Let now  $k': \mathcal{STS}(R, C \cup B_1) \rightarrow \mathcal{STS}(Q, B)$  be defined by the scheme

$(\gamma', (\lambda'_q)_{q \in Q}, (\lambda'_{c,b})_{c \in C \cup B_1, b \in B})$  defined as follows:

$$\begin{aligned} \gamma'(x) &= \left( \delta'(x) \wedge \bigwedge_{c \in C_1^S} \neg(x = c) \right) \vee \bigvee_{b \in B_1} (x = b) \\ \lambda'_q &= \psi'_q \text{ for each } q \in Q \\ \lambda'_{b,b} &= \text{true if } b \in B_1 \\ \lambda'_{c,b} &= \text{false if } b \in B_1 \text{ and } c \neq b \\ \lambda'_{c,b} &= \text{false if } b \in B_0 \text{ and } c \in C_1^S \\ \lambda'_{c,b} &= \bigvee_{h_0^{g(S)}(a)=b, h_0^S(d)=c} \kappa'_{d,a} \text{ if } b \in B_0 \text{ and } c \in C_0^S. \end{aligned}$$

It is now a routine verification that (for our fixed structure  $S$ )  $k'(h(S) \oplus B_1) = h(g(S))$ . Since all our definitions depend only on  $\zeta(S)$ , we also have  $k'(h(T) \oplus B_1) = h(g(T))$ .

One last step is required in this proof as the qfd operation  $k'$  may not preserve source separation for all structures, that is,  $k'$  may not lie in  $\mathcal{S}_{\text{sep}}$ . It does for the particular structures  $h(S) \oplus B_1$  and  $h(T) \oplus B_1$ , but perhaps not for others. Actually, structures  $U$  such that  $\zeta(U) \neq \zeta(h(S) \oplus B_1) = \zeta(h(T) \oplus B_1)$  do not matter in this context, so we can replace  $k'$  by the operation  $k$ , with the same domain and range as  $k'$ , which maps a structure  $U$  to  $k'(U)$  if  $\zeta(U) = \zeta(h(S) \oplus B_1)$ , and to the source-only source-separated structure  $B \in \mathcal{STS}(Q, B)$  where all relations are empty. This new operation  $k$  preserves source separation by construction, and it is easily verified to be qfd. This completes (at last) the proof.  $\square$

#### 4. The algebra $\mathcal{GP}$ of graphs with ports

Graphs with ports were introduced in Section 3.1. Recall that if  $P$  is a set of unary relation symbols, then  $E_P$  denotes the set  $E_P = \{\text{edge}\} \cup P$  and the class of graphs with ports in  $P$ , written  $\mathcal{GP}(P)$  can be identified with  $\mathcal{STS}(E_P)$ . We observe that a vertex of a graph with ports in  $P$  can be a  $p$ -port for one or several port labels  $p \in P$ , or for none at all.

For convenience, we will consider that  $P$  is a finite subset of the set  $\mathbb{N}$  of natural integers.

##### 4.1. The signature VR on graphs with ports

We define the set of sorts of the algebra  $\mathcal{GP}$  to be the set of finite subsets of  $\mathbb{N}$ . For each such subset  $P$ , the set of elements of  $\mathcal{GP}$  of sort  $P$  is the set  $\mathcal{GP}(P)$  of graphs with ports in  $P$ .

The signature VR consists of constants, unary operations and binary operations. These operations (interpreted in  $\mathcal{GP}$ ) are as follows.

First, if  $P, Q$  are finite subsets of  $\mathbb{N}$ , then  $\oplus$  is as in  $\mathcal{StS}$ , and is thus a binary operation of type  $(E_P, E_Q) \rightarrow E_{P \cup Q}$ . In  $\mathcal{GP}$ , we consider  $\oplus$  as an operation of type  $(P, Q) \rightarrow P \cup Q$ .

Next, the unary operations of VR are the following (clearly qfd) operations:

- if  $p, q$  are distinct integers,  $\text{add}_{p,q}$  is an operation of type  $P \rightarrow P$  for each sort  $P$  such that  $p, q \in P$ : it modifies neither the domain (the set of vertices) nor the unary relations  $p$  ( $p \in P$ ); the new edge relation has the existing edges, *plus* every edge from a  $p$ -port to a  $q$ -port: it is given by

$$\text{edge}(x, y) \vee (p(x) \wedge q(y));$$

- if  $D$  is a finite subset of  $\mathbb{N} \times \mathbb{N}$ ,  $\text{mdf}_D$  is an operation of type  $P \rightarrow Q$  where  $P$  is any finite set containing the domain of the relation  $D$  and  $Q$  is any finite set containing the range of  $D$ ; it modifies neither the domain (set of vertices) nor the edge relation; for each  $q \in Q$ , the  $q$ -ports of the output structure are the vertices of the input structure that are  $p$ -ports for some  $p$  such that  $(p, q) \in D$ ; that is,  $q(x)$  is given by  $\bigvee_{(p,q) \in D} p(x)$ .

Finally, for each integer  $p$ , we let  $\mathbf{p}$  be the constant of type  $\{p\}$  denoting the graph with a single vertex, no edges, and whose vertex is a  $p$ -port. We also let  $\mathbf{p}^{\text{loop}}$  be the same graph, with a single loop.

**Remark 4.1.** The following operations on graphs with ports occur in the literature, and are particular cases of VR-operations.

Let  $p \neq q$  be integers,  $P$  be a subset of  $\mathbb{N}$  containing  $p$  and  $Q = P \setminus \{p\} \cup \{q\}$ . The operation  $\text{ren}_{p \rightarrow q}$ , of type  $P \rightarrow Q$  which *renames* every  $p$ -port to a  $q$ -port, is an operation of VR: it is equal to  $\text{mdf}_D$  where  $D = \{(r, r) \mid r \in P \setminus \{p\}\} \cup \{(p, q)\}$ . Observe that this operation fuses the sets of vertices defined by  $p$  and  $q$ .

Let  $p$  be an integer, and let  $P$  be a subset of  $\mathbb{N}$  containing  $p$ . The operation  $\text{fg}_p$ , of type  $P \rightarrow P \setminus \{p\}$ , which *forgets*  $p$ -ports is an operation of VR: it is equal to  $\text{mdf}_D$  where  $D = \{(r, r) \mid r \in P \setminus \{p\}\}$ .

**Remark 4.2.** In our definition of graph with ports, an element of  $\mathcal{GP}(Q)$  does not need to have  $q$ -ports for each  $q \in Q$ . Thus, if  $P \subseteq Q$ , every graph with ports in  $P$  can also be viewed as a graph with ports in  $Q$ . The natural inclusion of  $\mathcal{GP}(P)$  into  $\mathcal{GP}(Q)$  is part of the signature VR: it is equal to  $\text{mdf}_D$  where  $D = \{(p, p) \mid p \in P\}$ .

**Remark 4.3.** Again (as in Example 3.4), the operations introduced in this section are denoted by overloaded symbols. A formal definition should specify the type of the operation, and would read something like  $\text{add}_{p,q,P}$  or  $\text{mdf}_{D,P,Q}$ . We prefer the more concise notation introduced here.

#### 4.2. A technical result

The following result describes the action of a qfd operation on a disjoint union of structures. It is the key to the main results of this section, described in Section 4.3 below.

**Proposition 4.4.** *Let  $\zeta$  be the type congruence (see Section 3.3). Let  $h$  be a unary qfd operation on  $\mathcal{StS}$ , from  $\mathcal{StS}(R, C)$  to  $\mathcal{StS}(E_Q, \emptyset) = \mathcal{GP}(Q)$ , let  $(R_1, C_1)$  and  $(R_2, C_2)$*

be sorts of  $StS$  such that  $R = R_1 \cup R_2$ ,  $C_1 \cap C_2 = \emptyset$  and  $C = C_1 \cup C_2$ , and let  $\vec{z} = (z_1, z_2)$  with  $z_1$  a  $\zeta$ -class in  $StS(R_1, C_1)$  and  $z_2$  a  $\zeta$ -class in  $StS(R_2, C_2)$ .

Then there exist quantifier-free definable operations  $g_{1,\vec{z}}: StS(R_1, C_1) \rightarrow \mathcal{GP}(Q_{1,\vec{z}})$ ,  $g_{2,\vec{z}}: StS(R_2, C_2) \rightarrow \mathcal{GP}(Q_{2,\vec{z}})$ , and  $f_{\vec{z}}: \mathcal{GP}(Q_{1,\vec{z}} \cup Q_{2,\vec{z}}) \rightarrow \mathcal{GP}(Q)$ , such that

- $f_{\vec{z}}$  is a composition of unary operations in  $VR$ ;
- for each  $x_1 \in StS(R_1, C_1)$  in class  $z_1$  and each  $x_2 \in StS(R_2, C_2)$  in class  $z_2$ ,  $h(x_1 \oplus x_2) = f_{\vec{z}}(g_{1,\vec{z}}(x_1) \oplus g_{2,\vec{z}}(x_2))$ .

**Proof.** Let  $(\delta, \psi_{\text{edge}}, (\psi_q)_{q \in Q})$  be the qfd operation scheme defining the operation  $h$ : here  $\psi_{\text{edge}}$  defines the edge relation,  $\psi_q$  defines the  $q$ -ports ( $q \in Q$ ), and there is no formula of the form  $\kappa_{c,d}$  since the range of  $h$  is in  $\mathcal{GP}(Q) = StS(E_Q, \emptyset)$ . The formulas  $\delta$ ,  $\psi_{\text{edge}}$  and  $\psi_q$ , for  $q \in Q$ , are in the language of  $(R, C)$ -structures.

The atoms of  $\delta(v)$  are either of the form  $r(y_1, \dots, y_{\rho(r)})$  ( $r \in R$ ), or  $v = c$ , or  $c_1 = c_2$  ( $c, c_1, c_2 \in C$ ). Let  $\delta^1$  be the formula obtained from  $\delta(v)$  by substituting the Boolean value 0 (false) for the following atoms, which are certainly false in a disjoint sum  $x_1 \oplus x_2$ , with  $x_1 \in StS(R_1, C_1)$ ,  $x_2 \in StS(R_2, C_2)$  and the variable  $v$  interpreted in  $x_1$ :

- each  $r$ -atom such that  $r \notin R_1$  and an argument of  $r$  is  $v$  or a constant in  $C_1$ ;
  - each  $r$ -atom such that  $r \notin R_2$  and an argument of  $r$  is a constant in  $C_2$ ;
  - each  $r$ -atom such that  $r \in R_1 \cap R_2$ , an argument of  $r$  is a constant in  $C_2$ , and another argument of  $r$  is  $v$  or a constant in  $C_1$ ;
  - each atom of the form  $y = c$  such that  $c \in C_2$  and  $y$  is equal to  $v$  or to a constant in  $C_1$ .
- The remaining atoms in  $\delta^1$  are either in  $QF(R_1, C_1, \{v\})$  or in  $QF(R_2, C_2, \emptyset)$ . Note that the  $\zeta$ -class of an element of  $StS(R_2, C_2)$  determines entirely which formulas in  $QF(R_2, C_2, \emptyset)$  it satisfies. For each  $\vec{z}$  as in the statement of the proposition, we let  $\delta^{1,\vec{z}}$  be the formula in  $QF(R_1, C_1, \{v\})$  obtained from  $\delta^1$  by replacing each atom in  $QF(R_2, C_2, \emptyset)$  by the Boolean value 0 or 1 according to the  $\zeta$ -class  $z_2$ . We observe that if  $v$  is a vertex of  $x_1 \oplus x_2$  which happens to be in  $x_1$ , then

$$\delta(v) \iff \delta^{1,\vec{z}}(v) \quad \text{whenever the } \zeta\text{-class of } x_2 \text{ is } z_2.$$

For each  $q \in Q$ , let  $\psi_q^{1,\vec{z}}$  be defined similarly. Then we also have, if  $v$  is a vertex of  $x_1 \oplus x_2$  in  $x_1$ ,

$$\psi_q(v) \iff \psi_q^{1,\vec{z}}(v) \quad \text{whenever the } \zeta\text{-class of } x_2 \text{ is } z_2.$$

Let also  $\delta^{2,\vec{z}}$  and  $\psi_q^{2,\vec{z}}$  be defined dually. And again, if  $i, j \in \{1, 2\}$ , we let  $\psi_{\text{edge}}^{i,j}(v, w)$  be the formula obtained from  $\psi_{\text{edge}}$  by substituting the Boolean value 0 for the atoms that are certainly false in a disjoint sum  $x_1 \oplus x_2$  for the variable  $v$  interpreted in  $x_i$  and the variable  $w$  interpreted in  $x_j$ :

- each  $r$ -atom such that  $r \notin R_i$  and  $v$  is an argument of  $r$ ;
- each  $r$ -atom such that  $r \notin R_j$  and  $w$  is an argument of  $r$ ;
- each  $r$ -atom such that  $r \notin R_1$  and a constant in  $C_1$  is an argument of  $r$ ;
- each  $r$ -atom such that  $r \notin R_2$  and a constant in  $C_2$  is an argument of  $r$ ;
- each  $r$ -atom such that  $r \in R_1 \cap R_2$ , an argument of  $r$  is a constant in  $C_2$ , and another argument of  $r$  is a constant in  $C_1$ ;

- each  $r$ -atom such that  $r \in R_1 \cap R_2$ , an argument of  $r$  is  $v$  (resp.  $w$ ) and another argument of  $r$  is a constant in  $C_{3-i}$  (resp.  $C_{3-j}$ );
  - each atom of the form  $v = c$  with  $c \in C_{3-i}$ ,  $w = c$  with  $c \in C_{3-j}$ , or  $c_1 = c_2$  with  $c_1 \in C_1$  and  $c_2 \in C_2$ ;
  - if  $i \neq j$ , each  $r$ -atom such that  $r \in R_1 \cap R_2$ , and  $v$  and  $w$  are arguments of  $r$ .
- As above, the remaining atoms in  $\psi_{\text{edge}}^{1,1}$  are in  $QF(R_1, C_1, \{v, w\}) \cup QF(R_2, C_2, \emptyset)$ , and for each  $\vec{z}$ , we let  $\psi_{\text{edge}}^{1,1,\vec{z}}$  be obtained from  $\psi_{\text{edge}}^{1,1}$  by substituting the Boolean values 0 or 1 for the atoms in  $QF(R_2, C_2, \emptyset)$  according to the  $\zeta$ -class  $z_2$ . If  $v, w$  are vertices of  $x_1 \oplus x_2$  in  $x_1$ , and if the  $\zeta$ -class of  $x_2$  is  $z_2$ , then

$$\psi_{\text{edge}}(v, w) \iff \psi_{\text{edge}}^{1,1,\vec{z}}(v, w).$$

We define  $\psi_{\text{edge}}^{2,2,\vec{z}}$  similarly, and get the analogous equivalence.

If  $i \neq j$ , the atoms of  $\psi_{\text{edge}}^{i,j}$  are in  $QF(R_i, C_i, \{v\})$  and in  $QF(R_j, C_j, \{w\})$ —which may include atoms in  $QF(R_1, C_1, \emptyset)$  and in  $QF(R_2, C_2, \emptyset)$ . Again, we let  $\psi_{\text{edge}}^{i,j,\vec{z}}$  be obtained from  $\psi_{\text{edge}}^{i,j}$  by substituting the Boolean values 0 or 1 for the atoms without free variables according to the  $\zeta$ -classes  $z_1$  and  $z_2$ . And we observe that if  $v, w$  are vertices of  $x_1 \oplus x_2$ ,  $v$  is in  $x_i$  and in the  $\zeta$ -class  $z_i$ ,  $w$  is in  $x_j$  and in the  $\zeta$ -class  $z_j$ , then

$$\psi_{\text{edge}}(v, w) \iff \psi_{\text{edge}}^{i,j,\vec{z}}(v, w).$$

Now let  $k = 1 + \max(Q)$ , let  $X_{k+1}, \dots, X_\ell$  be an enumeration of the subsets of  $QF(R_1, C_1, \{y\})$ , and let  $Y_{\ell+1}, \dots, Y_m$  be an enumeration of the subsets of  $QF(R_2, C_2, \{y\})$ . Let us denote by  $Q_1$  the set  $Q \cup \{k+1, \dots, \ell\}$  and by  $Q_2$  the set  $Q \cup \{\ell+1, \dots, m\}$ .

We define the qfd operation  $g_{1,\vec{z}}: StS(R_1, C_1) \rightarrow \mathcal{GP}(Q_1)$  defined by the following operation scheme:

$$\delta^{1,\vec{z}}, \quad \psi_{\text{edge}}^{1,1,\vec{z}}, \quad \psi_q^{1,\vec{z}} (q \in Q), \quad \theta_n (k+1 \leq n \leq \ell),$$

where for each  $k+1 \leq n \leq \ell$ ,  $\theta_n(v)$  holds if the set of quantifier-free formulas in  $QF(R_1, C_1, \{y\})$  satisfied by  $v$  is exactly  $X_n$ .

Similarly, the qfd operation  $g_{2,\vec{z}}: StS(R_2, C_2) \rightarrow \mathcal{GP}(Q_2)$  is defined by the operation scheme

$$\delta^{2,\vec{z}}, \quad \psi_{\text{edge}}^{2,2,\vec{z}}, \quad \psi_q^{2,\vec{z}} (q \in Q), \quad \theta_n (\ell+1 \leq n \leq m),$$

where for each  $\ell+1 \leq n \leq m$ ,  $\theta_n(v)$  holds if the set of quantifier-free formulas in  $QF(R_2, C_2, \{y\})$  satisfied by  $v$  is exactly  $X_n$ .

Finally, we consider structures  $x_1 \in StS(R_1, C_1)$  and  $x_2 \in StS(R_2, C_2)$ , with  $\zeta$ -classes respectively  $z_1$  and  $z_2$ , and we compare the graphs with ports  $g_{1,\vec{z}}(x_1) \oplus g_{2,\vec{z}}(x_2)$  and  $h(x_1 \oplus x_2)$ . The above remarks show that these two graphs have the same set of vertices, the same  $q$ -ports ( $q \in Q$ ), and the same edges between two vertices of  $x_1$  or two vertices of  $x_2$ . On the other hand,  $g_{1,\vec{z}}(x_1) \oplus g_{2,\vec{z}}(x_2)$  misses the edges of  $h(x_1 \oplus x_2)$  that connect a vertex of  $x_1$  with a vertex of  $x_2$ .

These edges are captured by the formulas  $\psi_{\text{edge}}^{1,2,\vec{z}}$  and  $\psi_{\text{edge}}^{2,1,\vec{z}}$ . Now, if  $v$  is a vertex of  $x_1$  and  $w$  is a vertex of  $x_2$ , we already observed that the truth values of  $\psi_{\text{edge}}^{1,2,\vec{z}}(v, w)$  and

$\psi_{\text{edge}}^{2,1,\vec{z}}(w, v)$  are entirely determined by the quantifier-free formulas with one free variable satisfied by  $v$  in  $x_1$  and by  $w$  in  $x_2$ : that is, they are entirely determined by the (unique) index  $k+1 \leq n \leq \ell$  such that  $\theta_n(v)$  and by the (unique) index  $\ell+1 \leq n \leq m$  such that  $\theta_n(w)$ . In other words,  $\psi_{\text{edge}}^{1,2,\vec{z}}(a, b)$  and  $\psi_{\text{edge}}^{2,1,\vec{z}}(b, a)$  are equivalent to disjunctions of conjunctions of the form

$$\theta_n(a) \wedge \theta_u(b) \quad \text{for some } k+1 \leq n \leq \ell \text{ and } \ell+1 \leq u \leq m.$$

Thus the edges in  $h(x_1 \oplus x_2)$  from a vertex of  $x_1$  to a vertex of  $x_2$  can be created from  $g_{1,\vec{z}}(x_1) \oplus g_{2,\vec{z}}(x_2)$  by applying repeatedly the operations (in VR) of the form  $\text{add}_{n,u}$  such that  $n \in [k+1, \ell]$ ,  $\theta_n \wedge \theta_u$  is a disjunct of  $\psi_{\text{edge}}^{1,2,\vec{z}}$ .

Similarly, the edges in  $h(x_1 \oplus x_2)$  from a vertex of  $x_2$  to a vertex of  $x_1$  can be created from  $g_{1,\vec{z}}(x_1) \oplus g_{2,\vec{z}}(x_2)$  by applying the appropriate operations of the form  $\text{add}_{u,n}$ . The last operation consists in forgetting the auxiliary ports numbered  $k+1$  to  $m$ , that is, in applying the operation  $\text{mdf}_D$ , with  $D = \{(q, q) \mid q \in Q\}$ .  $\square$

#### 4.3. Recognizable sets of graphs with ports

In this section, we consider different notions of recognizability that can be used for sets of graphs with ports. Let  $L \subseteq \mathcal{GP}(P)$ . Then  $L$  can be VR-recognizable, as a subset of the VR-algebra  $\mathcal{GP}$ . It can also be  $\mathcal{S}$ -recognizable, as a subset of the  $\mathcal{S}$ -algebra  $\text{StS}$  since  $\mathcal{GP}(P) = \text{StS}(E_P)$ . Finally, we introduce another signature, written  $\text{VR}^+$ , on  $\mathcal{GP}$ : it is obtained from VR by adding all the qfd operations between the sorts of  $\mathcal{GP}$ .

**Theorem 4.5.** *Let  $P$  be a finite subset of  $\mathbb{N}$  and let  $L$  be a subset of  $\mathcal{GP}(P)$ . The following properties are equivalent:*

1.  $L$  is  $\mathcal{S}$ -recognizable;
2.  $L$  is  $\text{VR}^+$ -recognizable;
3.  $L$  is VR-recognizable;

**Proof.** Since the operations of VR are operations of  $\text{VR}^+$ , and the operations of  $\text{VR}^+$  are operations of  $\mathcal{S}$ , it follows from Proposition 2.1 that (1) implies (2), and (2) implies (3). Thus, we only need to verify that (3) implies (1).

We use Lemma 2.5, with  $\mathcal{F} = \text{VR}$ ,  $\mathcal{S} = \mathcal{GP}$ ,  $\mathcal{G} = \mathcal{S}$ ,  $T = \text{StS}$ , and  $\zeta$  the type congruence (see Section 3.3), which relates structures with sources of the same sort, provided they satisfy the same quantifier-free formulas. We use the collection  $\mathcal{H}$  of sets  $\mathcal{H}_{(R,C),P}$  of unary qfd operations from  $\text{StS}(R, C)$  to  $\mathcal{GP}(P)$ .

Let  $L$  be a VR-recognizable subset of  $\mathcal{GP}(P)$  and let  $\equiv$  be a locally finite VR-congruence on  $\mathcal{GP}$  such that  $L$  is a union of  $\equiv$ -classes. Since  $\zeta$  is a locally finite  $\mathcal{S}$ -congruence on  $\text{StS}$  (Proposition 3.9), its restriction to  $\mathcal{GP}$  is also a locally finite VR-congruence; and the intersection of  $\equiv$  and  $\zeta$  is a locally finite VR-congruence on  $\mathcal{GP}$  which saturates  $L$ . Thus we can assume, without loss of generality, that  $\equiv$ -equivalent elements of  $\mathcal{GP}$  are also  $\zeta$ -equivalent.

Next we consider the equivalence relation  $\approx$  on  $\text{StS}$  defined as in Lemma 2.5. Note that the identity of  $\mathcal{GP}(P)$  belongs to  $\mathcal{H}_{(E_P, \emptyset), P}$ , so that  $\approx$ -equivalent elements of  $\mathcal{GP}(P) = \text{StS}(E_P, \emptyset)$  are also  $\equiv$ -equivalent. In particular,  $\approx$  saturates  $L$  and it suffices to show that

$\approx$  is locally finite and is a  $\mathcal{S}$ -congruence. In view of Lemma 2.5, it is enough to verify that  $\mathcal{H}$  satisfies the assumptions of Lemmas 2.3 and 2.4.

We first verify the hypothesis of Lemma 2.3. Let  $g$  be an operation of  $\mathcal{S}$ : either  $g$  is a unary qfd operation or  $g = \oplus$ . In the latter case, Proposition 4.4 states precisely that the required property holds.

If  $g$  is a qfd operation of type  $(R_1, C_1) \rightarrow (R, C)$ , and  $h \in \mathcal{H}_{(R,C),P}$ , then  $h \circ g$  is a qfd operation (Lemma 3.7) and hence,  $h_1 = h \circ g \in \mathcal{H}_{(R_1,C_1),P}$ . Now letting  $f$  be the identity mapping of  $\mathcal{GP}(P)$ , we find that  $h(g(x)) = f(h_1(x))$  as required. In this case,  $h_1$  and  $f$  do not depend on the  $\zeta$ -class of  $x$ .

Next, we turn to the verification of the hypothesis of Lemma 2.4. Let  $\varphi_1, \dots, \varphi_k$  be an enumeration of the elements of  $QF(R, C, \{x\})$  and let  $\chi_1, \dots, \chi_\ell$  be an enumeration of the elements of  $QF(R, C, \{x, y\})$ .

Thus, a qfd operation scheme from  $StS(R, C)$  into  $\mathcal{GP}(Q)$  consists in the choice of a formula  $\delta = \varphi_{i_0}$  ( $1 \leq i_0 \leq k$ ), a formula  $\psi_{\text{edge}} = \chi_j$  ( $1 \leq j \leq \ell$ ), a sequence of formulas  $\varphi_{i_1}, \dots, \varphi_{i_r}$  ( $1 \leq i_1 < \dots < i_r \leq k$ ), and a partition of  $Q$  as  $Q = Q_1 \cup \dots \cup Q_r$ : if  $q \in Q_j$ , then  $\psi_q = \varphi_{i_j}$ . (If  $Q = \emptyset$ , then  $r = 0$ .)

Let us now consider two unary qfd operations  $g: StS(R, C) \rightarrow \mathcal{GP}(Q)$  and  $g': StS(R, C) \rightarrow \mathcal{GP}(Q')$ , associated with the same choice of values  $i_0, j$  and  $i_1 < \dots < i_r$ . Let  $Q = Q_1 \cup \dots \cup Q_r$  and  $Q' = Q'_1 \cup \dots \cup Q'_r$  be the corresponding partitions of  $Q$  and  $Q'$ . Finally let  $\pi, \pi_0, \pi_1, \dots, \pi_r$  be the following operations in the signature VR. These operations have the common particularity to not alter the graph structure, and to modify only the port predicates.

The mapping  $\pi_0$  shifts every port index of an element of  $\mathcal{GP}(Q)$  by  $m = \max(Q')$ , to yield a graph with ports in  $Q+m$ , whose port names do not intersect  $Q'$ . We let  $R_h = Q_h + m$  for  $1 \leq h \leq r$ .

For  $1 \leq h \leq r$ ,  $\pi_h = \text{mdf}_{D_h}$ , where

$$D_h = \left\{ (a, a) \mid a \in \bigcup_{i < h} Q'_i \cup \bigcup_{i > h} R_i \right\} \cup (R_h \times Q'_h).$$

Thus  $\pi_h$  turns a graph with ports in  $Q'_1 + \dots + Q'_{h-1} + R_h + R_{h+1} + \dots + R_r$  into a graph with ports in  $Q'_1 + \dots + Q'_{h-1} + Q'_h + R_{h+1} + \dots + R_r$ , with the same vertex set, the same edge relation, the same  $q$ -ports for each  $q \in \bigcup_{i < h} Q'_i \cup \bigcup_{i > h} R_i$ , and with each  $r$ -port ( $r \in R_h$ ) turned into a  $q$ -port for each  $q \in Q'_h$ .

It is now an easy verification that, if  $\pi = \pi_r \circ \dots \circ \pi_1 \circ \pi_0$ , then  $g'(x) = \pi(g(x))$  for each  $x \in StS(R, C)$ . Thus the quasi-order  $\leq_{(R,C)}$  defined in Lemma 2.4 is in fact a finite index equivalence relation, and this concludes the proof.  $\square$

**Remark 4.6.** This actually proves also that we get the same recognizable sets of graphs with ports, if we consider  $\mathcal{GP}(Q)$  as a domain of sort  $Q$  in the algebra of structures *without* sources—which consists of the domains  $StS(R, \emptyset)$  equipped with the operations of  $\mathcal{S}$  between them. If we were only interested in the equivalence of this recognizability with VR- and VR<sup>+</sup>-recognizability (or just the equivalence between VR- and VR<sup>+</sup>-recognizability), we could do with Lemmas 2.3 and 2.4 instead of Lemma 2.5, and with a simpler version of Proposition 4.4, making no reference to  $\zeta$ .

#### 4.4. Variants of the algebra of graphs with ports

The first variant considered here replaces the signature VR by a smaller signature, which we will see is equivalent to VR in terms of recognizability. The second one concerns a certain class of graphs with ports, and is central in the definition of the clique-width of a finite graph.

##### 4.4.1. A variant of VR on $\mathcal{GP}$

In Section 4.3, we exhibited signatures larger than VR, for which all the VR-recognizable sets of graphs with ports are recognizable: namely the signature  $\text{VR}^+$  on  $\mathcal{GP}$  and the signature  $\mathcal{S}$  on the wider algebra  $\text{St}\mathcal{S}$ . In contrast, we exhibit in this section a smaller signature (in fact, a signature consisting of VR-derived operations) which does not create new recognizable subsets.

The basic idea behind the definition of this new signature is the following: when we evaluate a VR-term  $t$  of the form  $\text{add}_{p,q}(t')$ , then we add edges from each  $p$ -port of  $G'$ , the value of  $t'$ , to each of its  $q$ -ports. It may happen that some edges from a  $p$ -port to a  $q$ -port already exist in  $G'$ . In this case, we do not add a parallel edge since we are dealing with simple graphs. Thus the term  $t$  presents a form of redundancy, since some of its edges may be, in some sense, defined twice.

For disjoint sets of port labels  $P$  and  $Q$ , we denote by  $J(P, Q)$  the set of VR-derived unary operations defined by terms of the form  $f_1(f_2(\dots(f_n(x))\dots))$ , where the  $f_i$  are of the forms  $\text{add}_{p,q}$  or  $\text{add}_{q,p}$  for  $p$  in  $P$  and  $q$  in  $Q$ . Since the operations  $\text{add}_{p,q}$  are idempotent and commute with one another, an operation in  $J(P, Q)$  is completely described by a subset of  $(P \times Q) \cup (Q \times P)$ . Thus  $J(P, Q)$  is finite, although one can write infinitely many terms specifying its elements. For each element  $J \in J(P, Q)$ , we let  $\otimes_J$  denote the binary operation defined, for  $G \in \mathcal{GP}(P)$  and  $H \in \mathcal{GP}(Q)$ , by  $G \otimes_J H = J(G \oplus H)$ .

We observe that in the evaluation of a term of the form  $t \otimes_J t'$ , the application of  $\otimes_J$  does not recreate edges that already exist in  $G$ , the value of  $t$ , or in  $G'$ , the value of  $t'$  since the  $\text{add}_{p,q}$  operations forming  $\otimes_J$  add edges between the disjoint graphs  $G$  and  $G'$  (because  $p$  and  $q$  are not port labels of the same argument graphs).

Now the signature NLC consists of the operations  $\otimes_J$  as above, the unary qfd operations of the form  $\text{fg}_p$  and  $\text{ren}_{p \rightarrow q}$  as defined in Remark 4.1, and the constants  $\mathbf{p}$  and  $\mathbf{p}^{\text{loop}}$  as in VR. We denote by  $\mathcal{GP}^{\text{NLC}}$  the NLC-algebra of graphs with ports.

**Remark 4.7.** The notation NLC refers to a very similar algebra used by Wanke [44].

**Example 4.8.** We have in fact already encountered NLC-operations and NLC-derived operations.

The VR-derived operation  $f_z$  whose existence is proved in Proposition 4.4 is actually NLC-derived. Consider indeed the last paragraphs of the proof of that proposition: the operation  $f_z$  is obtained by first composing operations of the form  $\text{add}_{n,u}$  and  $\text{add}_{u,n}$ , where the pairs  $(n, u)$  lie in a certain subset of  $[k+1, \ell] \times [\ell+1, m]$  and the pairs  $(u, n)$  lie in another subset of  $[\ell+1, m] \times [k+1, \ell]$ , and then composing operations of the form  $\text{fg}_p$ .

One can also check that the operations  $\pi_0, \dots, \pi_r$  at the end of the proof of Theorem 4.5 are NLC-derived.

**Proposition 4.9.** *Let  $P$  be a finite subset of  $\mathbb{N}$  and let  $L$  be a subset of  $\mathcal{GP}(P)$ . Then  $L$  is VR-recognizable if and only if  $L$  is NLC-recognizable.*

**Proof.** The proof is a simple extension of the proof of Theorem 4.5.

Since the operations of NLC are VR-derived, every VR-recognizable subset of  $\mathcal{GP}$  is NLC-recognizable. For the converse, we observe that the proof that (1) implies (3) in Theorem 4.5 can be modified to show that an NLC-recognizable set of  $\mathcal{GP}$  is  $\mathcal{S}$ -recognizable.

Again, we rely on Lemma 2.5, but now with  $\mathcal{F} = \text{NLC}$ ,  $\mathcal{S} = \mathcal{GP}$ , and  $\mathcal{G}, T, \zeta$  and  $\mathcal{H}$  as in Theorem 4.5.

In order to justify the fact that the arguments used in the proof of Theorem 4.5 are also valid with these assumptions, we refer to Example 4.8. Indeed this example shows two things: on the one hand, the operation  $f_z$  in Proposition 4.4 is in fact NLC-derived, so that the first hypothesis of Lemma 2.5 is satisfied by this new choice of  $\mathcal{F}$  and  $\mathcal{S}$ . On the other hand, the finiteness hypothesis of Lemma 2.4 is also satisfied with this new value of  $\mathcal{F} = \text{NLC}$ . This completes the proof.  $\square$

#### 4.4.2. Graphs whose port labels partition the vertex set

In certain contexts, and in particular in the definition of the clique-width of a graph (see Remark 4.11), one needs to consider graphs with ports where port labels partition the vertex set. More precisely, for each set of port labels  $P$ , let  $\mathcal{GP}^\pi(P)$  be the set of elements of  $\mathcal{GP}(P)$  such that each vertex is a port, and no vertex is both a  $p$ -port and a  $q$ -port for  $p \neq q$ . Let also  $\mathcal{GP}^\pi = (\mathcal{GP}^\pi(P))$ .

Note that  $\mathcal{GP}^\pi$  is preserved by the operations of the form  $\oplus$ ,  $\text{add}_{p,q}$  and  $\text{ren}_{p \rightarrow q}$ . These operations form the signature  $\text{VR}^\pi$ , and  $\mathcal{GP}^\pi$  is a  $\text{VR}^\pi$ -algebra.

**Remark 4.10.** The operation  $\text{add}_{p,q}$  is written  $\alpha_{p,q}$  in [19].

**Remark 4.11.** The *clique-width* of a finite graph  $G$ , denoted by  $\text{cwd}(G)$ , is defined as the smallest cardinality of a set  $P$  such that  $G$  is the value of a (finite)  $\text{VR}^\pi$ -term using a set  $P$  of port labels, see [19,7].

For algorithmic applications [20], it is useful to have efficient recognition algorithms for classes of graphs of clique-width at most  $k$ . At the moment we only know that this problem is NP. It is polynomial for  $k \leq 3$ , see [7].

**Proposition 4.12.** *Let  $L$  be a subset of  $\mathcal{GP}^\pi(P)$ . Then  $L$  is  $\text{VR}^\pi$ -recognizable if and only if  $L$  is VR-recognizable.*

**Proof.** Since  $\text{VR}^\pi$  consists of operations in VR, every locally finite VR-congruence on  $\mathcal{GP}$  induces a locally finite  $\text{VR}^\pi$ -congruence on  $\mathcal{GP}^\pi$ . In particular, if  $L$  is VR-recognizable, and hence is saturated by a locally finite VR-congruence on  $\mathcal{GP}$ , then  $L$  is saturated by a locally finite  $\text{VR}^\pi$ -congruence on  $\mathcal{GP}^\pi$ , and hence  $L$  is  $\text{VR}^\pi$ -recognizable.

To prove the converse, we first introduce the mapping  $\sigma: \mathcal{GP} \rightarrow \mathcal{GP}^\pi$  defined as follows. If  $G \in \mathcal{GP}(P)$ , then  $\sigma(G)$  is the graph in  $\mathcal{GP}^\pi(2^P)$  with the same set of vertices and the same edge relation as  $G$ , and such that for each vertex  $v$  and each  $X \subseteq P$ ,  $v$  is an  $X$ -port in  $\sigma(G)$  if and only if  $X$  is the set of  $p \in P$  such that  $v$  is a  $p$ -port in  $G$ . We say that a port label  $p$  is *void* in  $G$  if there are no  $p$ -ports in  $G$ .

Now let us assume that  $L$  is  $\text{VR}^\pi$ -recognizable, and let  $\equiv$  be a locally finite congruence on  $\mathcal{GP}^\pi$  saturating it. If  $G, H \in \mathcal{GP}(P)$ , we let  $G \sim H$  if  $\sigma(G)$  and  $\sigma(H)$  have the same non-void port labels, and  $\sigma(G) \equiv \sigma(H)$ . It is immediately verified that  $\sim$  is a locally finite equivalence relation.

We now verify that  $\sim$  is a VR-congruence. If  $G \in \mathcal{GP}(P)$  and  $H \in \mathcal{GP}(Q)$ , it is easily seen that  $\sigma(G \oplus H) = \sigma(G) \oplus \sigma(H)$ . If  $p, q \in P$ , then  $\sigma(\text{add}_{p,q}(G)) = f(\sigma(G))$  where  $f$  is the composition of the operations  $\text{add}_{X,Y}$  for each  $X, Y \subseteq P$  such that  $p \in X$  and  $q \in Y$ . Finally, one can verify that if  $D \subseteq P \times Q$ , then  $\sigma(\text{mdf}_D(G)) = g(\sigma(G))$  where  $g$  is the composition of the operations  $\text{ren}_{X \rightarrow Y}$ , where  $X \subseteq P, Y \subseteq Q$  and  $Y = D^{-1}(X) = \{q \in Q \mid (p, q) \in D \text{ for some } p \in P\}$ .

It is a routine task to derive from these observations the fact that  $\sim$  is a VR-congruence. We now need to verify that  $\sim$  saturates  $L$ . Let  $G \in L$  and  $G \sim H$ . In particular,  $G \in \mathcal{GP}^\pi$ , so that the non-void port labels of  $\sigma(G)$  are exactly the sets  $\{p\}$  where  $p$  is a non-void port label of  $G$ . Since  $\sigma(G)$  and  $\sigma(H)$  have the same non-void port labels,  $H$  is also in  $\mathcal{GP}^\pi$ . Moreover, if  $h$  is the composition of the operations  $\text{ren}_{\{p\} \rightarrow p}$  ( $p$  non-void in  $G$ ), then  $G = h(\sigma(G))$  and  $H = h(\sigma(H))$ . Since  $h$  is  $\text{VR}^\pi$ -derived, it follows that  $G \equiv H$ , and hence  $H \in L$ . This concludes the proof.  $\square$

## 5. The algebra of graphs with sources

Recall that we call *graphs with sources* the elements of  $\text{StS}$  of sort  $(E, C)$ , where  $E = \{\text{edge}\}$  and  $C$  is some finite set of source labels, and that we write  $\mathcal{GS}(C)$  for  $\text{StS}(E, C)$  (see Section 3.1).

### 5.1. The signature HR

The disjoint union  $\oplus$  and the operations of the form  $\text{srcren}_{a \rightarrow b}$ ,  $\text{srcfg}_a$  and  $\text{fus}_{a,b}$  (defined in Example 3.4) preserve graphs with sources. We denote by HR the signature consisting of all these operations, so  $\mathcal{GS}$  is an HR-algebra.

We note the following properties of HR-recognizability.

**Proposition 5.1.** *Let  $C$  be a finite set of source labels. Every  $\mathcal{S}$ -recognizable subset of  $\text{StS}(E, C)$  is HR-recognizable.*

**Proof.** This is a simple consequence of Proposition 2.1 and of the observation given above that the operations of HR are also operations of  $\mathcal{S}$ .  $\square$

Note that the class **Graph** of graphs, defined in Section 3.1, is equal to  $\mathcal{GP}(\emptyset)$  as well as to  $\mathcal{GS}(\emptyset) = \text{StS}(E)$ . Thus VR-recognizability and HR-recognizability are properties of subsets of **Graph**.

**Corollary 5.2.** *Let  $L$  be a set of graphs (a subset of **Graph**). If  $L$  is VR-recognizable, then it is HR-recognizable.*

**Proof.** This follows immediately from Proposition 5.1 and Theorem 4.5.  $\square$

**Remark 5.3.** Intuitively, the VR-operations are more powerful than the HR-operations (every HR-context-free set of simple graphs is VR-context-free but the converse is not true, [12]), but the HR-operations are not among the VR-operations, nor are they derived from them.

We will see in Sections 6.1 and 6.2 sufficient conditions for HR-recognizable sets to be VR-recognizable, and in Section 6.3, examples of HR-recognizable sets which are not VR-recognizable.

## 5.2. Variants of the algebra of graphs with sources

We find in the literature a number of variants of the signature HR or of the algebra  $\mathcal{GS}$ . We now discuss these different variants, to verify that they do not introduce artefacts from the point of view of recognizability.

### 5.2.1. The signature $\text{HR}_{\parallel}$

Let  $\text{HR}_{\parallel}$  denote the signature on  $\mathcal{GS}$  obtained by substituting the parallel composition  $\parallel$  for  $\oplus$  (see Section 3.5.1). With the same proof as Proposition 3.15, we get the following result.

**Proposition 5.4.** *Let  $L$  be a subset of  $\mathcal{GS}$ . Then  $L$  is HR-recognizable if and only if it is  $\text{HR}_{\parallel}$ -recognizable.*

### 5.2.2. Source-separated graphs

As in Section 3.5.2, we now discuss the class  $\mathcal{GS}_{\text{sep}}$  of source separated graphs. The operations of HR all preserve source separation, except for  $\text{fus}_{a,b}$ , but we defined in Example 3.16 the operation  $\text{fus}_{a \rightarrow b} = \text{srcfg}_a \circ \text{fus}_{a,b}$  which does. Let  $\text{HR}_{\text{sep}}$  be the signature on  $\mathcal{GS}_{\text{sep}}$  consisting of  $\oplus$  and the qfd unary operations of the form  $\text{srcren}_{a \rightarrow b}$ ,  $\text{srcfg}_a$  and  $\text{fus}_{a \rightarrow b}$ .

**Proposition 5.5.** *Let  $L$  be a subset of  $\mathcal{GS}_{\text{sep}}$ . Then  $L$  is HR-recognizable if and only if it is  $\text{HR}_{\text{sep}}$ -recognizable.*

**Proof.** Since  $\text{HR}_{\text{sep}}$  consists only of HR-derived operations, every HR-recognizable set subset of  $\mathcal{GS}_{\text{sep}}$  is also  $\text{HR}_{\text{sep}}$ -recognizable.

The proof of the converse is a variant of the proof of Theorem 3.18. First we note that the type relation  $\zeta$  (see Section 3.3) is also an HR-congruence on  $\mathcal{GS}$ . We use the same mapping  $h$  defined in the proof of Theorem 3.18, that maps a graph with sources  $S \in \mathcal{GS}(C)$  to a source-separated graph  $h(S) \in \mathcal{GS}_{\text{sep}}(C)$  by splitting sources that were identified in  $S$ . We refer to that proof for notation used here.

If  $L$  is an  $\text{HR}_{\text{sep}}$ -recognizable subset of  $\mathcal{GS}_{\text{sep}}$  and  $\equiv$  is a locally finite  $\text{HR}_{\text{sep}}$ -congruence recognizing it, we define a relation  $\sim$  on  $\mathcal{GS}$  as follows. If  $S, T \in \mathcal{GS}(C)$ , we say that  $S \sim T$  if  $\zeta(S) = \zeta(T)$  and  $h(S) \equiv h(T)$ . As in the proof of Theorem 3.18,  $\sim$  is easily seen

to be a locally finite equivalence relation. It is also easily seen that  $\sim$  is preserved under the  $\text{HR}_{\text{sep}}$ -operation  $\oplus$ .

We now need to verify that if  $S \sim T \in \mathcal{GS}(C)$  and  $g$  is one of the unary operations of  $\text{HR}_{\text{sep}}$  defined on  $\mathcal{GS}(C)$ , then  $g(S) \sim g(T)$ . Again, Proposition 3.9 shows that  $\zeta(g(\zeta(S))) = \zeta(g(\zeta(T)))$  and we want to show that  $h(g(S)) \equiv h(g(T))$ . The graphs  $S$  and  $T$  are fixed for the rest of this proof. We write  $h_0$ ,  $C_0$  and  $C_1$  for  $h_0^S$ ,  $C_0^S$  and  $C_1^S$ .

As in the proof of Theorem 3.18, it suffices to construct an  $\text{HR}_{\text{sep}}$ -derived operation  $k$ , depending on  $g$  and  $\zeta(S)$ , such that  $h(g(S)) = k(h(S))$  and  $h(g(T)) = k(h(T))$ . There is no reason why the operation  $k$  constructed in the proof of Theorem 3.18 should be  $\text{HR}_{\text{sep}}$ -derived, but the operations  $g$  considered here, namely  $\text{srcr}_{a \rightarrow b}$ ,  $\text{srcfg}_a$  and  $\text{fus}_{a \rightarrow b}$  are simple enough that we can directly construct a suitable  $k$  in each case.

If  $g = \text{srcr}_{a \rightarrow b}$ . Then  $g$  is defined on  $\mathcal{GS}(C)$  (where  $a \in C$  and  $b \notin C$ ) and its range is  $\mathcal{GS}(C \setminus \{a\} \cup \{b\})$ . One verifies that  $h(\text{srcr}_{a \rightarrow b}(S))$  is equal to

- $\text{srcr}_{a \rightarrow b}(h(S))$  if  $a \in C_1$  and  $b > h_0(a)$ ;
- $\text{srcr}_{a \rightarrow h_0(a)}(\text{srcr}_{h_0(a) \rightarrow b}(h(S)))$  if  $a \in C_1$  and  $b < h_0(a)$ ;
- $\text{srcr}_{a \rightarrow b}(h(S))$  if  $a \in C_0$  and  $b < c$  for every  $c \in C_1$  such that  $h_0(c) = a$ ;
- $\text{srcr}_{c \rightarrow b}(\text{srcr}_{b \rightarrow c}(h(S)))$  if  $a \in C_0$  and  $b > c = \min\{d \in C_1 \mid h_0(d) = a\}$ .

If  $g = \text{srcfg}_a$ . Then  $g$  is defined on  $\mathcal{GS}(C)$  (where  $a \in C$ ) and its range is  $\mathcal{GS}(C \setminus a)$ . One verifies that  $h(\text{srcfg}_a(S))$  is equal to

- $\text{fus}_{a \rightarrow h_0(a)}(h(S))$  if  $a \in C_1$ ;
- $\text{fus}_{a \rightarrow c}$  if  $a \in C_0$ ,  $h_0^{-1}(a) \neq \emptyset$  and  $c = \min\{h_0^{-1}(a)\}$ ;
- $\text{srcfg}_a(h(S))$  if  $a \in C_0$  and  $h_0^{-1}(a) = \emptyset$ .

If  $g = \text{fus}_{a \rightarrow b}$ . Then  $g$  is defined on  $\mathcal{GS}(C)$  (where  $a \neq b \in C$ ) and its range is  $\mathcal{GS}(C \setminus a)$ . One verifies that  $h(\text{fus}_{a \rightarrow b}(S))$  is equal to

- $\text{srcr}_{a \rightarrow h_0(a)}(\text{fus}_{h_0(a) \rightarrow h_0(b)}(h(S)))$  if  $a \in C_1$  and  $h_0(b) < h_0(a)$ ;
- $\text{srcr}_{a \rightarrow h_0(b)}(\text{fus}_{h_0(b) \rightarrow h_0(a)}(h(S)))$  if  $a \in C_1$  and  $h_0(b) > h_0(a)$ ;
- $\text{srcr}_{a \rightarrow h_0(a)}(h(S))$  if  $a \in C_1$  and  $h_0(b) = h_0(a)$ ;
- $\text{fus}_{a \rightarrow h_0(b)}(h(S))$  if  $a \in C_0$  and  $a > h_0(b)$ ;
- $\text{srcr}_{a \rightarrow c}(\text{fus}_{h_0(b) \rightarrow a}(h(S)))$  if  $a \in C_0$ , and  $c = \min\{h_0(b), h_0^{-1}(a)\}$ , and  $a < h_0(b)$ ;
- $\text{srcr}_{a \rightarrow c}(\text{fus}_{a \rightarrow c}(h(S)))$  if  $a \in C_0$ ,  $a = h_0(b)$  and  $c = \min\{d \in C_1 \mid h_0(d) = a\}$ .

This concludes the proof.  $\square$

Again with the same proof as for Proposition 3.15, we can show that the operation  $\oplus$  can be replaced by  $\parallel$  in the signature  $\text{HR}_{\text{sep}}$ —yielding the signature  $\text{HR}_{\text{sep}, \parallel}$ .

**Proposition 5.6.** *Let  $L$  be a subset of  $\mathcal{GS}_{\text{sep}}$ . Then  $L$  is  $\text{HR}_{\text{sep}}$ -recognizable if and only if it is  $\text{HR}_{\text{sep}, \parallel}$ -recognizable.*

### 5.2.3. Other variants

The equivalence between  $\text{HR}_{\text{sep}, \parallel}$ - and  $\text{HR}_{\parallel}$ -recognizability for a set of source-separated graphs—a consequence of Propositions 5.4–5.6—was already established by Courcelle [10] for graphs with multi-edges (see Section 7). In the same paper, Courcelle established the equivalence between  $\text{HR}_{\text{sep}}$ - and  $\mathcal{B}$ -recognizability for several variants  $\mathcal{B}$  of the signature  $\text{HR}$ , which we now describe. We refer to [10] for the proofs.

For each finite set  $C$  of source labels, let  $\text{srcfg}_{all}$  be the composition of the operations  $\text{srcfg}_c$  for each  $c \in C$  (in any order). Let also  $\square_C$  be the following binary operation on  $\mathcal{GS}_{\text{sep}}$ , of type  $(C, C) \rightarrow \emptyset$ : if  $G, H \in \mathcal{GS}_{\text{sep}}(C)$ , then  $G \square_C H = \text{srcfg}_{all}(G \parallel H)$ :  $G \square_C H$  is obtained by first taking the parallel composition  $G \parallel H$ , and then forgetting all source labels.

Let  $\mathcal{CS}$  be the signature on  $\mathcal{GS}_{\text{sep}}$ , which consists only of the  $\square_C$  operations.

Let  $\text{HR}^{\text{fg}}$  be the derived signature of  $\text{HR}_{\parallel}$ , which consists of the operations  $\text{srcfg}_{all}$  and  $\parallel$ .

Let  $\text{HR}^{\text{ren}}$  be the subsignature of  $\text{HR}_{\parallel}$ , which consists of the operations  $\text{srcren}_{p \rightarrow q}$  and  $\parallel$ .

Let  $\text{HR}_{\text{sep}}^{\text{ren}}$  be the subsignature of  $\text{HR}_{\text{sep}, \parallel}$ , which consists of the operations  $\parallel$  and those operations  $\text{srcren}_{p \rightarrow q}$  which preserve source separation.

The following result is a compilation of [10, Section 4].

**Proposition 5.7.** *If  $L \subseteq \mathcal{GS}$ , then  $L$  is HR-recognizable if and only if  $L$  is  $\text{HR}^{\text{ren}}$ -recognizable.*

*If  $L \subseteq \mathcal{GS}_{\text{sep}}$ , then  $L$  is  $\text{HR}_{\text{sep}}$ -recognizable if and only if  $L$  is  $\text{HR}_{\text{sep}}^{\text{ren}}$ -recognizable.*

*If  $L \subseteq \text{Graph}$ , the following are equivalent:*

- $L$  is HR-recognizable;
- $L$  is  $\mathcal{CS}$ -recognizable;
- $L$  is  $\text{HR}^{\text{fg}}$ -recognizable.

**Remark 5.8.** The notation  $\mathcal{CS}$  refers to the notion of *fully cutset-regular* sets of graphs, introduced by Abrahamson and Fellows [1]. Full cutset-regularity is equivalent to  $\mathcal{CS}$ -recognizability.

In [10], Courcelle also shows a number of closure properties of the class of  $\text{HR}_{\text{sep}}$ -recognizable sets of source-separated graphs with sources. In particular, it is shown that this class contains all singletons and it is closed under the operations of  $\text{HR}_{\text{sep}}$  [10, Section 6].

Finally Courcelle shows the following result [10, Theorem 6.7].

**Proposition 5.9.** *Let  $L \in \mathcal{GS}(C)$ . Then  $L$  is HR-recognizable if and only if  $\text{srcfg}_{all}(L)$  is HR-recognizable.*

## 6. Finiteness conditions ensuring that HR- and VR-recognizability coincide

We saw that a VR-recognizable set of graphs is always HR-recognizable (Corollary 5.2). The converse does not hold in general, as we discuss in Section 6.3. We first explore structural conditions on graphs, which are sufficient to guarantee that an HR-recognizable set of graphs is also VR-recognizable.

Let  $\vec{K}_{n,n}$  be the directed complete bipartite graph with  $n + n$  vertices. A directed graph  $G \in \text{Graph}$  is *without*  $\vec{K}_{n,n}$  if it has no subgraph isomorphic to  $\vec{K}_{n,n}$ . The main result in this section is the following.

**Theorem 6.1.** *Let  $n$  be an integer. An HR-recognizable set of graphs without  $\vec{K}_{n,n}$  is VR-recognizable.*

This theorem is proved in Section 6.1, and some of its corollaries are discussed in Section 6.2.

Note that results similar to Corollary 5.2 and Theorem 6.1 hold for VR- and HR-equational sets of graphs. As explained in the introduction, such sets are exactly the context-free sets of graphs, formally specified in terms of recursive sets of equations using the operations of VR and HR respectively. Specifically, the following results are known to hold:

- every HR-equational set of simple directed graphs is VR-equational [15];
- if a VR-equational set of directed graphs is without  $\vec{K}_{n,n}$  for some  $n$ , then it is HR-equational (by the main theorem in [12] and Lemma 6.6).

Thus the same combinatorial condition is sufficient to guarantee the equivalence between VR- and HR-recognizability, as well as between VR- and HR-equationality. A further similar result concerning monadic second-order definability and using a stronger combinatorial property will be discussed in Section 6.4.

### 6.1. Proof of Theorem 6.1

We first record the following observation.

**Lemma 6.2.** *Let  $G$  be a directed graph and let  $x, y$  be two vertices of  $G$  that are not adjacent, and such that there is no vertex  $z$  such that both  $(x, z)$  and  $(y, z)$  (resp. both  $(z, x)$  and  $(z, y)$ ) are edges. Let  $H$  be obtained from  $G$  by identifying  $x$  and  $y$ . If  $G$  contains  $\vec{K}_{m,m}$  as a subgraph, then so does  $H$ .*

**Proof.** Let  $K$  be a subgraph of  $G$  isomorphic to  $\vec{K}_{m,m}$ . From the hypothesis, the vertices  $x$  and  $y$  are not both in  $K$ . It follows that  $K$  is still isomorphic to a subgraph of  $H$ .  $\square$

The proof of Theorem 6.1 will proceed as follows. We consider an HR-recognizable set  $L$  of finite graphs without  $\vec{K}_{n,n}$  and we denote by  $m$  the largest integer such that  $\vec{K}_{m,m}$  is a subgraph of a graph in  $L$ . Such an integer exists by hypothesis.

Since we are talking about source-less graphs, the set  $L$  is  $\text{HR}_{\text{sep}}$ -recognizable by Proposition 5.5, and we consider a locally finite  $\text{HR}_{\text{sep}}$ -congruence  $\equiv$  saturating  $L$ . We will define a locally finite NLC-congruence  $\sim$  on  $\mathcal{GP}$  that also saturates  $L$ . By Proposition 4.9, this suffices to show that  $L$  is VR-recognizable. The definition of  $\sim$  makes use of the notion of expansion of a graph, defined below.

Note that the following definitions depend on the integer  $m$ , even though terminology and notation do not make this dependence explicit.

*Small and large port labels and formulas.* Let  $G \in \mathcal{GP}(P)$  be a graph with ports. If  $p \in P$ , we denote by  $p_G$  the set of  $p$ -ports of  $G$ . We say that a port label  $p$  is *void* in  $G$  if  $p_G$  is empty, we say that  $p$  is *small* in  $G$  if  $1 \leq \text{card}(p_G) \leq m$  and that  $p$  is *large* in  $G$  if  $\text{card}(p_G) > m$ .

Observe that if the port labels  $p$  and  $q$  are both large in  $G$ , then  $\text{add}_{p,q}(G)$  contains  $\vec{K}_{m+1,m+1}$  as a subgraph.

Moreover, if  $p$  is large in  $G$ , if  $r_1, \dots, r_k$  are small in  $G$ , let

$$H = \text{add}_{p,r_1} \text{add}_{p,r_2} \cdots \text{add}_{p,r_k}(G).$$

For  $i = 1, \dots, k$ , let  $n_i = \text{card}(r_i G)$ . If  $H$  does not contain  $\vec{K}_{m+1,m+1}$ , then we must have  $n_1 + \dots + n_k \leq m$ . If  $G$  already contains edges from the  $p$ -ports to other vertices, then  $n_1 + \dots + n_k < m$ . The notion of expansion below will make it possible to handle this sort of complicated situation (see Example 6.3).

Let us say that a closed first-order formula is *small* if it has quantifier-depth at most  $2m + 2$ . Note that the existence of a subgraph isomorphic to  $\vec{K}_{m+1,m+1}$  can be expressed by a first-order formula of quantifier-depth  $2m + 2$ .

**Expansions.** We will define supergraphs of  $G \in \mathcal{GP}(P)$  called *expansions*, that contain information relevant to the distribution of small and large port labels, and where ports are represented by sources. Furthermore, it will be possible to simulate an NLC-operation on  $G$  that does not create  $\vec{K}_{m+1,m+1}$  subgraphs by HR-operations on expansions of  $G$ . These expansions will then be used to transform the  $\text{HR}_{\text{sep}}$ -congruence  $\equiv$  into an NLC-congruence  $\sim$ .

Furthermore, we will define  $\sim$  in such a way that two equivalent graphs satisfy the same small first-order formulas.

We now give formal definitions. For each port label  $p$ , we define a set  $C(p)$  of source labels,

$$C(p) = \{\text{in}(p, i), \text{out}(p, i), s(p, i) \mid 1 \leq i \leq m\}.$$

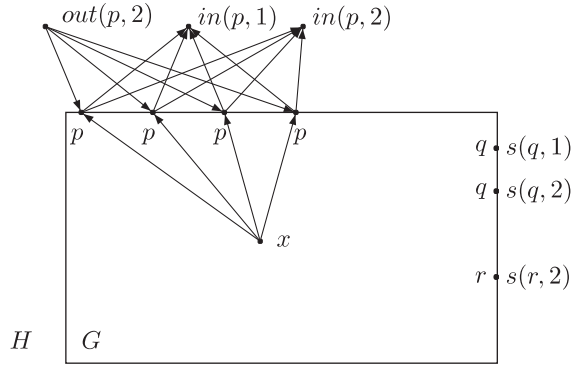
If  $P$  is a set of port labels,  $C(P)$  denotes the union of the  $C(p)$ , for  $p$  in  $P$ .

Let  $G \in \mathcal{GP}(P)$  be a graph with ports, let  $C \subseteq C(P)$ , and let  $\bar{G}$  be a graph in  $\mathcal{GS}_{\text{sep}}(C)$ . We say that  $\bar{G}$  is an *expansion* of  $G$  if the following conditions hold:

- (1)  $\bar{G}$  has no subgraph isomorphic to  $\vec{K}_{m+1,m+1}$ .
- (2) Except for the labeling of ports and sources,  $G$  is a subgraph of  $\bar{G}$ . The sources of  $\bar{G}$ , and its vertices and edges not in  $G$ , are specified by Conditions (3) and (4).
- (3) If  $p$  is small in  $G$ , then each  $p$ -port of  $G$  is an  $s(p, i)$ -source of  $\bar{G}$  for some integer  $i \leq m$ . Different  $p$ -ports are of course labelled by different source labels. There are no  $\text{in}(p, j)$ - or  $\text{out}(p, j)$ -sources.
- (4) If  $p$  is large in  $G$ , then there may be vertices of  $\bar{G}$  that are not in  $G$ , with source labels of the form  $\text{in}(p, i)$  or  $\text{out}(p, i)$  for some  $i \leq m$ . Moreover, there is an edge in  $\bar{G}$  from each vertex of  $p_G$  to each  $\text{in}(p, i)$ -source, and from each  $\text{out}(p, i)$ -source to each vertex in  $p_G$ . There are no  $s(p, j)$ -sources.

In particular,  $G$  may have several different expansions, but it has only a finite number of expansions (up to isomorphism). This number is bounded by a function depending on  $m$  and the cardinality of  $P$ . Indeed, for each small port label  $p$ , there is only a bounded number of ways to make  $p$ -ports into  $s(p, i)$ -sources (see (3)), and for each large port label  $p$ , there is a bounded number of ways to create  $\text{in}(p, i)$ - and  $\text{out}(p, i)$ -sources (see (4)).

**Example 6.3.** Let  $m = 2$ , and let  $G$  be a graph with port labels  $p, q, r$ . Suppose that  $G$  has 4  $p$ -ports, 2  $q$ -ports and 1  $r$ -port, so that  $p$  is large, and  $q, r$  are small in  $G$ , see Fig. 1. Then in any expansion of  $G$ , every  $q$ - and  $r$ -port will be a source, say labeled by  $s(q, 1), s(q, 2)$

Fig. 1.  $H$  is an expansion of  $G$ .

and  $s(r, 2)$  (there is only one  $s(r, i)$ -source, but it is not required that these sources should be labeled with consecutive numbers starting at 1).

Moreover, an expansion of  $G$  may have up to two new vertices that are  $in(p, j)$ -sources, and at most one  $out(p, j)$ -source. Say, an expansion  $H$  could have new vertices as  $in(p, 1)$ - and  $in(p, 2)$ -sources, with edges from each of the 4  $p$ -ports to each  $in(p, j)$ -source; and it could have a new vertex as a, say,  $out(p, 2)$ -source, with edges from that vertex to each of the  $p$ -ports.

Note that if  $G$  has a vertex  $x$  with an edge from  $x$  to at least 3  $p$ -sources, then an expansion cannot have 2  $out(p, j)$ -sources: otherwise it would contain a copy of  $\vec{K}_{3,3}$ , which is not allowed for an expansion.

**Remark 6.4.** It is not always the case that  $G$  is determined by each of its expansions  $\tilde{G}$ . If  $p$  is large in  $G$  but  $\tilde{G}$  has no  $in(p, i)$ - or  $out(p, i)$ -sources, then it is not possible to determine which of its vertices are  $p$ -ports.

*Construction of an NLC-congruence from an  $HR_{\text{sep}}$ -congruence.* Let  $\equiv$  be a locally finite  $HR_{\text{sep}}$ -congruence saturating  $L$ . We define a relation  $\sim$  on  $\mathcal{GP}$  as follows. For  $G$  and  $G'$  in  $\mathcal{GP}(P)$  we let  $G \sim G'$  if and only if

- (a) either  $G$  and  $G'$  both contain  $\vec{K}_{m+1, m+1}$  as a subgraph, or neither does and in that case, the following two conditions hold:
- (b)  $G$  and  $G'$  satisfy the same small first-order formulas (i.e., with quantifier-depth at most  $2m + 2$ ) on graphs with ports.
- (c) for every expansion  $\tilde{G}$  of  $G$ , there exists an expansion  $\tilde{G}'$  of  $G'$  such that  $\tilde{G} \equiv \tilde{G}'$  and  $\tilde{G}$  and  $\tilde{G}'$  satisfy the same small first-order formulas on graphs with sources (we say that  $\tilde{G}$  and  $\tilde{G}'$  are *equivalent expansions*); and conversely, for every expansion  $\tilde{G}'$  of  $G'$  there exists an expansion  $\tilde{G}$  of  $G$  equivalent to  $\tilde{G}'$ .

Note that Condition (b) implies that  $G$  and  $G'$  have the same void, small and large port labels, and Condition (c) implies that  $\tilde{G}$  and  $\tilde{G}'$  have the same source labels.

The relation  $\sim$  is clearly an equivalence relation on each set  $\mathcal{GP}(P)$ . It has finitely many classes on each  $\mathcal{GP}(P)$  since a finite graph has a uniformly bounded number of expansions

(up to isomorphism), the  $\text{HR}_{\text{sep}}$ -congruence  $\equiv$  is locally finite, and there are finitely many first-order formulas of each quantifier-depth on graphs with sources in a subset of  $C(P)$ .

Now a graph without ports and without  $\vec{K}_{m+1,m+1}$  has a unique expansion: itself. It follows that, for graphs without ports and without  $\vec{K}_{m+1,m+1}$ , the equivalences  $\equiv$  and  $\sim$  coincide. In particular,  $\sim$  saturates  $L$  since  $\equiv$  does.

It remains to prove that  $\sim$  is an NLC-congruence. Recall that the signature NLC consists of the operations of the form  $\text{fg}_p$ ,  $\text{ren}_{p \rightarrow q}$  and  $\otimes_J$ .

*The port forgetting operation.* We first consider the operation  $\text{fg}_p$ . We consider  $G, G'$  with  $G \sim G'$  and we want to prove that  $H \sim H'$ , where  $H = \text{fg}_p(G)$  and  $H' = \text{fg}_p(G')$ .

First of all, the underlying graphs of  $G$  and  $H$  (resp.  $G'$  and  $H'$ ) are identical, so that  $G$  and  $G'$  contain  $\vec{K}_{m+1,m+1}$  if and only if so do  $H$  and  $H'$ . If this is the case, then  $G \sim G'$  and  $H \sim H'$ . We now exclude this case and assume that  $G$  and  $G'$  are without  $\vec{K}_{m+1,m+1}$ . Note also that if  $p$  is void in  $G$ , then it is in  $G'$  as well, and we have  $H = G$ ,  $H' = G'$ , so that  $H \sim H'$ . We now assume that  $p$  is not void in  $G$ .

It is an immediate consequence of Theorem 3.12 that  $H$  and  $H'$  satisfy the same small first-order formulas on graphs with ports, so Condition (b) is verified.

We now consider Condition (c). Let  $\tilde{H}$  be an expansion of  $H$ . We will show that there exists an expansion  $\tilde{G}$  of  $G$  and a unary  $\text{HR}_{\text{sep}}$ -term  $t$  such that  $\tilde{H} = t(\tilde{G})$ . Since  $G \sim G'$ , there exists an equivalent expansion  $\tilde{G}'$  of  $G'$ , and  $t(\tilde{G}')$  will be the desired expansion of  $H'$ . Using the fact that  $\equiv$  is an  $\text{HR}_{\text{sep}}$ -congruence and Theorem 3.12, we will have  $H \sim H'$  as expected.

If  $p$  is large in  $G$ , the situation is particularly simple:  $\tilde{H}$  is also an expansion of  $G$ , so we can choose  $t$  to represent the identity. If  $\tilde{G}'$  is an expansion of  $G'$ , equivalent to  $\tilde{H}$ , then  $\tilde{G}'$  does not use source labels of the form  $s(p, i)$ ,  $\text{in}(p, i)$  or  $\text{out}(p, i)$ , so  $\tilde{G}'$  is also an expansion of  $H'$ .

If  $p$  is small in  $G$ , let  $\tilde{G}$  be a graph with source obtained from  $\tilde{H}$  by letting each  $p$ -port of  $G$  be an  $s(p, i)$ -source (where distinct source labels are used for distinct  $p$ -ports). Then  $\tilde{G}$  is an expansion of  $G$ , and  $\tilde{H} = t(\tilde{G})$  where  $t$  is the composition of the operations  $\text{srcfg}_{s(p, i)}$  ( $1 \leq i \leq m$ ). Using the definition of  $\sim$ , there exists an expansion  $\tilde{G}'$  of  $G'$  which is equivalent to  $\tilde{G}$ , and we only need to verify that  $\tilde{H}' = t(\tilde{G}')$  is an expansion of  $H'$ . The only point to check here is the fact that  $H'$  is a subgraph of  $\tilde{H}'$ : this follows from the facts that  $G$  is a subgraph of  $\tilde{G}$  and the operations  $t$  and  $\text{fg}_p$  do not change the underlying graph structures.

*The renaming operation.* We now consider the operation  $\text{ren}_{p \rightarrow q}$ . Let  $G, G'$  with  $G \sim G'$ : as with the port forgetting operation, we want to prove that  $H \sim H'$  where  $H = \text{ren}_{p \rightarrow q}(G)$  and  $H' = \text{ren}_{p \rightarrow q}(G')$ . As above, we can reduce the proof to the case where neither  $G$  nor  $G'$  contains  $\vec{K}_{m+1,m+1}$ , and where  $p$  is not void in  $G$  (if  $p$  is void in  $G$ , then  $H = G$  and  $H' = G'$ ). Moreover, Condition (b) follows from Theorem 3.12.

We consider Condition (c), following the same strategy as above. Let  $\tilde{H}$  be an expansion of  $H$ .

If  $q$  is void in  $G$ , then the transformation  $\text{ren}_{p \rightarrow q}$  is a reversible renaming, that is,  $G = \text{ren}_{q \rightarrow p}(H)$ . Moreover, if  $t$  is the composition of the operations of the form  $\text{srcren}_{s(p, i) \rightarrow s(q, i)}$ ,  $\text{srcren}_{\text{in}(p, i) \rightarrow \text{in}(q, i)}$  and  $\text{srcren}_{\text{out}(p, i) \rightarrow \text{out}(q, i)}$ , and

if  $t'$  is the composition of the operations  $\text{srcren}_{s(q,i) \rightarrow s(p,i)}$ ,  $\text{srcren}_{in(q,i) \rightarrow in(p,i)}$  and  $\text{srcren}_{out(q,i) \rightarrow out(p,i)}$ , then  $\bar{G} = t'(\bar{H})$  is an expansion of  $G$ ,  $\bar{H} = t(\bar{G})$ . Moreover, if  $\bar{G}'$  is an expansion of  $G'$ , equivalent to  $\bar{G}$ , then  $\bar{H}' = t(\bar{G}')$  is an expansion of  $H'$ .

We now assume that  $q$  is not void in  $G$ . We need to consider several cases.

*Case 1:*  $p$  and  $q$  are both large in  $G$ . Then  $p$  is void and  $q$  is large in  $H$ .

In order to build the desired  $\bar{G}$ , we split each  $in(q, i)$ -source of  $\bar{H}$  into an  $in(p, i)$ -source and an  $in(q, i)$ -source. The  $in(p, i)$ -source is linked by incoming edges to all  $p$ -ports of  $G$ , and the  $in(q, i)$ -source is linked similarly to all  $q$ -ports. In the same fashion, we split each  $out(q, i)$ -source of  $\bar{H}$  into an  $out(p, i)$ -source and an  $out(q, i)$ -source linked by outgoing edges to all  $p$ -ports of  $G$  and to all  $q$ -ports respectively. The term  $t$  such that  $\bar{H} = t(\bar{G})$  is the composition of the operations  $\text{fus}_{in(p,i) \rightarrow in(q,i)}$  and  $\text{fus}_{out(p,i) \rightarrow out(q,i)}$ .

The graph  $\bar{G}$  does not contain  $\bar{K}_{m+1, m+1}$ , since  $\bar{H}$  does not (by Lemma 6.2). Hence  $\bar{G}$  is an expansion of  $G$ . Let now  $\bar{G}'$  be an expansion of  $G'$  equivalent to  $\bar{G}$ , and let  $\bar{H}' = t(\bar{G}')$ . It is easily verified that  $\bar{H}'$  is an expansion of  $H'$ , and as above, it follows that  $H \sim H'$ .

*Case 2:*  $p$  is small and  $q$  is large in  $G$ .

In order to build  $\bar{G}$  from  $\bar{H}$ , we make the  $p$ -ports of  $G$  into  $s(p, i)$ -sources, we delete the edges between the  $in(q, i)$ - and the  $out(q, i)$ -sources and the  $p$ -ports of  $G$ . The term  $t$  which must do the opposite (that is, construct  $\bar{H}$  from  $\bar{G}$ ) is a composition of source forgetting operations and of additions of new edges. More precisely, for each  $i, j$  such that  $s(p, i)$  and  $in(q, j)$  are source labels in  $\bar{G}$ , we use the operation  $Z \mapsto Z \oplus (\alpha \rightarrow \omega)$ , where  $(\alpha \rightarrow \omega)$  is the 2-vertex, 2-source, 1-edge graph, followed by the operations  $\text{fus}_{\alpha \rightarrow s(p,i)}$  and  $\text{fus}_{\omega \rightarrow in(q,j)}$ . We then apply similar operations to create edges from the  $out(q, j)$ - to the  $s(p, i)$ -sources. And we finally apply the operations  $\text{srcfg}_{s(p,i)}$ .

The graph  $\bar{G}$  is a subgraph of  $\bar{H}$  (up to source labels), so  $\bar{G}$  does not contain  $\bar{K}_{m+1, m+1}$ , and hence it is an expansion of  $G$ . The proof continues as in the previous case.

*Case 3:*  $q$  is small and  $p$  is large in  $G$ .

To build  $\bar{G}$  from  $\bar{H}$ , we make the  $q$ -ports of  $G$  into  $s(q, i)$ -sources, we delete the edges between the  $in(p, i)$ -sources or the  $out(p, i)$ -sources and the  $q$ -ports of  $G$ . In addition we rename each  $in(p, i)$ -source to an  $in(q, i)$ -source, and each  $out(p, i)$ -source to an  $out(q, i)$ -source. We can use the same reasoning as in Case 2 to conclude in this case.

*Case 4:*  $p$  and  $q$  are small in  $G$ , and  $\text{card}(p_G) + \text{card}(q_G) \leq m$ .

To build  $\bar{G}$  from  $\bar{H}$ , we rename  $s(q, i)$  into  $s(p, i)$  whenever the  $s(q, i)$ -source of  $\bar{H}$  is a  $p$ -port in  $G$ . The term  $t$  which does the opposite is a composition of source renamings. The graph  $\bar{G}$  does not contain  $\bar{K}_{m+1, m+1}$ , otherwise  $\bar{H}$  would do, since  $\bar{G}$  is equal to  $\bar{H}$  up to source labels, and hence  $\bar{G}$  is an expansion of  $G$ . The other parts of the proof are the same.

*Case 5:*  $p$  and  $q$  are small in  $G$ , and  $\text{card}(p_G) + \text{card}(q_G) \geq m + 1$ .

To build  $\bar{G}$  from  $\bar{H}$ , we make the  $p$ -ports (resp.  $q$ -ports) of  $G$  into  $s(p, i)$ -sources (resp.  $s(q, i)$ -sources), we delete the edges between the  $in(q, i)$ - and  $out(q, i)$ -sources and the  $p$ - and  $q$ -ports of  $G$ , and we delete the  $in(q, i)$ - and  $out(q, i)$ -sources. The term  $t$  which does the opposite is a composition of additions of new edges and of  $\text{srcfg}$  operations, as in Case 2, see Fig. 2. The graph  $\bar{G}$  does not contain  $\bar{K}_{m+1, m+1}$ , otherwise  $\bar{H}$  would too, since  $\bar{G}$  is a subgraph of  $\bar{H}$  (up to source labels), and hence  $\bar{G}$  is an expansion of  $G$ . The proof continues as in the previous cases.

This concludes the proof that  $G \sim G'$  implies  $\text{ren}_{p \rightarrow q}(G) \sim \text{ren}_{p \rightarrow q}(G')$ .

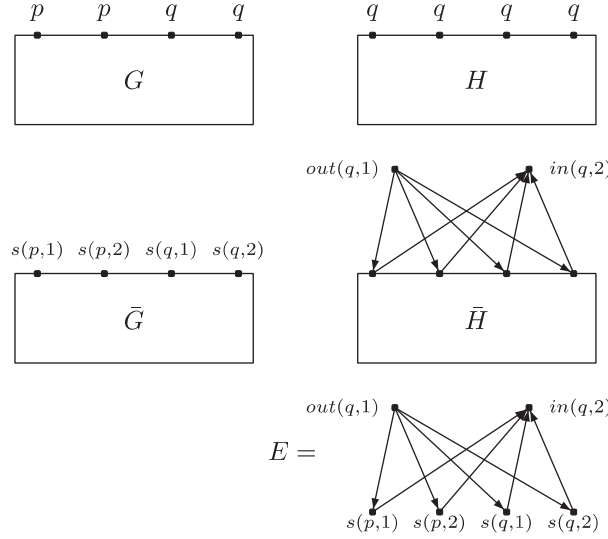


Fig. 2.  $m = 2$  and  $\bar{H} = t(\bar{G}) = \text{srcfg}_{s(p,1),s(p,2),s(q,1),s(q,2)}(\bar{G} \parallel E)$ .

*The operation  $\otimes_J$ .* We now consider the operation  $\otimes_J$  where  $J \subseteq (P \times Q) \cup (Q \times P)$ ,  $P$  and  $Q$  are disjoint. Let  $G \sim G'$  in  $\mathcal{GP}(P)$ ,  $K \sim K'$  in  $\mathcal{GP}(Q)$ ,  $H = G \otimes_J K$  and  $H' = G' \otimes_J K'$ . We want to prove that  $H \sim H'$ .

We first consider the very special case where  $J = \emptyset$ , and the operation  $\otimes_J$  is simply the disjoint union. Then  $H$  contains  $\vec{K}_{m+1,m+1}$  if and only if  $G$  or  $K$  does, if and only if  $G'$  or  $K'$  does, if and only if  $H'$  does.

Assuming that  $H$  does not contain  $\vec{K}_{m+1,m+1}$ , an application of Theorem 3.12 ensures, as for the operations of port forgetting or renaming that  $H$  and  $H'$  satisfy the same small first-order formulas.

We now consider an expansion  $\bar{H}$  of  $H$ . It is necessarily of the form  $\bar{H} = \bar{G} \oplus \bar{K}$  where  $\bar{G}$  and  $\bar{K}$  are expansions of  $G$  and  $K$  respectively. Then there exist expansions  $\bar{G}'$  and  $\bar{K}'$  of  $G'$  and  $K'$  respectively, which are equivalent to  $\bar{G}$  and  $\bar{K}$ . One then verifies that  $\bar{H}' = \bar{G}' \oplus \bar{K}'$  is an expansion of  $H'$ , which is equivalent to  $\bar{H}$ .

Next we assume that  $J$  is a singleton,  $J = \{(p, q)\}$ , that is,  $G \otimes_J K = \text{add}_{p,q}(G \oplus K)$  with  $p \in P$  and  $q \in Q$ .

Since  $G$  and  $G'$  on one hand, and  $K$  and  $K'$  on the other satisfy the same small first-order formulas, Theorem 3.12 shows that  $H = \text{add}_{p,q}(G \oplus K)$  contains  $\vec{K}_{m+1,m+1}$  if and only if  $H' = \text{add}_{p,q}(G' \oplus K')$  does. Assume now this is not the case and consider an expansion  $\bar{H}$  of  $H$ .

Again there are several cases. Note that  $p$  and  $q$  cannot both be large in  $G$  and  $K$  respectively. We claim that  $\bar{H}$  can be defined as  $t(\bar{G}, \bar{K})$  where  $t$  is an  $\text{HR}_{\text{sep}}$ -term,  $\bar{G}$  is an expansion of  $G$  and  $\bar{K}$  is an expansion of  $K$ . As for the other operations, we consider expansions  $\bar{G}'$  and  $\bar{K}'$  of  $G'$  and  $K'$ , equivalent to  $\bar{G}$  and  $\bar{K}$ . Although it is a bit tedious, we verify formally that  $\bar{H}' = t(\bar{G}' \oplus \bar{K}')$  is an expansion of  $H'$ . It follows that  $\bar{H}'$  is equivalent to  $\bar{H}$ , and hence  $H \sim H'$ .

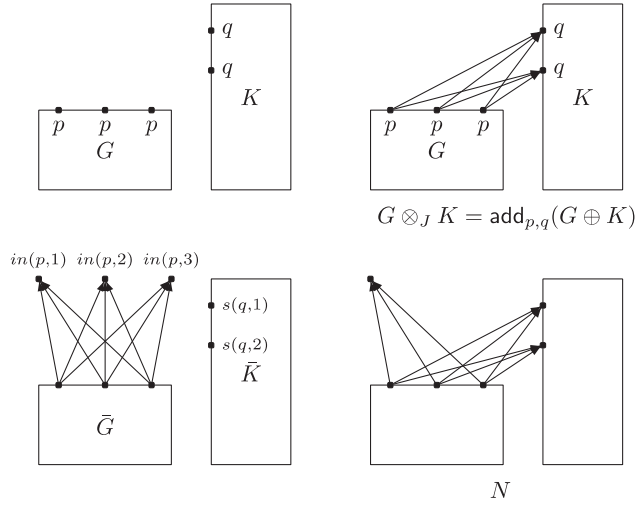


Fig. 3.  $N = t(\bar{G}, \bar{K}) = \text{srcfg}_{all}(\bar{G} \parallel \text{srcren}_{s(q,1) \rightarrow in(p,2), s(q,2) \rightarrow in(p,3)}(\bar{K}))$ .

*Case 1:*  $p$  is large in  $G$  and  $q$  is small in  $K$ .

Then  $H$  has edges from all  $p$ -ports of  $G$  to all  $q$ -ports of  $K$ , which are actually  $s(q, i)$ -sources in  $\bar{H}$ . For each of these  $s(q, i)$ -sources, say  $x$ , we create a new vertex  $x'$ , and each edge coming from  $G$  to  $x$  is redirected towards  $x'$ . We make  $x'$  into an  $in(p, j)$ -source (for some appropriate  $j$ ) of the expansion  $\bar{G}$  of  $G$  we are constructing. The desired expansion  $\bar{K}$  of  $K$  is just the subgraph of  $\bar{H}$  induced by the set of vertices of  $K$ . And  $\bar{G}$  consists of the subgraph of  $\bar{H}$  induced by the vertices of  $G$  together with  $x'$  and all these redirected edges. Then the  $\text{HR}_{\text{sep}}$ -term  $t$  needs only to fuse in  $\bar{G} \oplus \bar{K}$  the above described  $in(p, j)$ -sources with the corresponding  $s(q, i)$ -sources. This can be done by a combination of the operation  $\oplus$  and those of the form  $\text{fus}_{in(p, j) \rightarrow s(q, i)}$ . The only point to check is that  $\bar{G}$  does not contain  $\bar{K}_{m+1, m+1}$ . We can apply Lemma 6.2 because  $\bar{H}$  is obtained from  $\bar{G} \oplus \bar{K}$  by fusions of pairs of vertices which are not adjacent and have no incoming edges with the same source (because  $G$  and  $K$  are disjoint) and no outgoing edge at all.

Then there exist expansions  $\bar{G}'$  and  $\bar{K}'$  of  $G'$  and  $K'$  respectively, equivalent to  $\bar{G}$  and  $\bar{K}$ . By letting  $\bar{H}' = t(\bar{G}', \bar{K}')$ , we get the desired expansion of  $H'$ , equivalent to  $\bar{H}$ .

This case is illustrated in Fig. 3, where  $m = 3$  and  $N$  is the constructed expansion of  $G \otimes_J K$ .

*Case 2:*  $p$  is small in  $G$  and  $q$  is large in  $K$ .

It is fully similar to the first case, creating new  $out(q, j)$ -sources instead of  $in(p, j)$ -sources. We omit the details.

*Case 3:*  $p$  is small in  $G$  and  $q$  is small in  $K$ .

Let  $\bar{G}$  be the subgraph with sources of  $\bar{H}$  consisting of the vertices of  $G$ , and let  $\bar{K}$  be defined similarly in terms of  $K$ . Then  $\bar{H}$  is obtained from  $\bar{G} \oplus \bar{K}$  by the addition of edges from each  $s(p, i)$ -source of  $\bar{G}$  to each  $s(q, j)$ -source of  $\bar{K}$ , which can be done by an  $\text{HR}_{\text{sep}}$ -term (see Case 2 of the discussion of the renaming operation). Since  $\bar{G}$  and  $\bar{K}$  are

subgraphs of  $\bar{H}$ , they cannot contain  $\vec{K}_{m+1,m+1}$  and hence, they are in fact expansions of  $G$  and  $K$  as desired. The proof continues as above.

*Case 4:*  $p$  is void in  $G$  or  $q$  is void in  $K$ .

Then  $\text{add}_{p,r}$  acts as the identity on  $G \oplus K$ , so  $\otimes_J$  acts as  $\oplus$  on  $(G, K)$  and we are back to a previously studied case. Recall that if  $p$  (resp.  $q$ ) is void in  $G$  (resp.  $K$ ), then it is void in every  $\sim$ -equivalent graph with source.

This concludes the study of the case where  $J$  is a singleton in  $P \times Q$ . The case where  $J$  is a singleton in  $Q \times P$  is of course similar.

The proof is actually the same in the general case where  $J$  is not a singleton. We need only do the same constructions for all elements  $(p, q)$  in  $J$ . The only possible difficulty could arise from the use of Lemma 6.2 to verify that the graphs  $\bar{G}$  and  $\bar{K}$  obtained from  $\bar{H}$  by the creation of vertices (like  $x'$  in Case 1 above) and the redirection of edges do not contain  $\vec{K}_{m+1,m+1}$ , and hence are expansions. Thus let us consider the transformation of  $\bar{G} \oplus \bar{K}$  into  $\bar{H}$ . It consists in a sequence of fusions of pairs of vertices. Whenever we fuse an  $\text{in}(p, i)$ -source of  $\bar{G}$ , say  $x$ , with an  $\text{in}(q, j)$ -source of  $\bar{K}$ , say  $y$ , we must verify that the fusions performed previously keep the hypothesis of Lemma 6.2 valid. It is clear that  $x$  and  $y$  are not adjacent, since  $x$  is adjacent with vertices of  $G$  only. Because of previous fusions, there may exist an edge from some  $z$  in  $G$  to  $y$ . However, this edge comes from a previously applied operation  $\text{add}_{p',q}$  with  $p' \neq p$ . It follows that there is no edge from  $z$  to  $x$ . An analogous argument also applies to fusions between an  $\text{out}(p, i)$ -source of  $G$  and an  $\text{in}(q, j)$ -source of  $K$ , and also when we exchange the roles of  $G$  and  $K$ . Hence, finally, we can apply Lemma 6.2 to deduce that  $\bar{G}$  and  $\bar{K}$  do not contain  $\vec{K}_{m+1,m+1}$  because  $\bar{H}$  does not. Hence, they are expansions of  $G$  and  $K$ , as we needed to check.

This concludes the proof of Theorem 6.1.

## 6.2. Other finiteness conditions

We now consider some consequences of Theorem 6.1. Let  $K_{n,n}$  be the undirected complete bipartite graph with  $n + n$  vertices, that is,  $K_{n,n}$  is the undirected graph underlying  $\vec{K}_{n,n}$ . We say that a (directed) graph is *without*  $K_{n,n}$  if its undirected underlying graph has no subgraph isomorphic to  $K_{n,n}$ .

We say that a graph  $G$  is *uniformly  $k$ -sparse* if  $\text{card}(E(H)) \leq k \text{ card}(V(H))$  for every finite subgraph  $H$  of  $G$ , where  $V(H)$  and  $E(H)$  are the sets of vertices and edges of  $H$ . A set of graphs is *uniformly  $k$ -sparse* if each of its elements is.

**Proposition 6.5.** *Let  $L \subseteq \text{Graph}$  be a set of graphs, satisfying one of the following properties:*

- $L$  is without  $\vec{K}_{n,n}$  for some  $n$*
  - or  $L$  is without  $K_{n,n}$  for some  $n$*
  - or  $L$  consists only of planar graphs*
  - or  $L$  is uniformly  $k$ -sparse for some  $k$*
  - or  $L$  consists only of graphs of tree-width at most  $k$  for some  $k$ .*
- Then  $L$  is HR-recognizable if and only if  $L$  is VR-recognizable.*

**Proof.** By Corollary 5.2, it is always the case that a VR-recognizable set of graphs is HR-recognizable.

If  $L$  is without  $\vec{K}_{n,n}$  for some  $n$ , the converse implication was proved in Theorem 6.1. Lemma 6.6 below shows that  $L$  is without  $K_{p,p}$  for some  $p$  if and only if it is without  $\vec{K}_{n,n}$  for some  $n$ .

It is well-known that planar graphs are without  $K_{3,3}$  (planarity is a property of the underlying undirected graph, and  $K_{3,3}$  is the undirected graph underlying  $\vec{K}_{3,3}$ ). It follows that planar graphs are also without  $\vec{K}_{3,3}$ , and the result follows from Theorem 6.1.

It is easily seen that  $\vec{K}_{2k+1,2k+1}$  is not  $k$ -sparse. So if  $L$  is uniformly  $k$ -sparse, then it is without  $\vec{K}_{2k+1,2k+1}$ .

Finally, it is known that graphs of tree-width at most  $k$  are uniformly  $(k+1)$ -sparse (see for instance [16]), which yields the last assertion.  $\square$

**Lemma 6.6.** *Let  $p$  be an integer. There exists an integer  $n$  such that a directed graph without  $\vec{K}_{p,p}$ , is without  $K_{n,n}$ .*

**Proof.** We use the particular case of Ramsey's Theorem for bipartite graphs, given as Theorem 1 in [27, p. 95]. It states that for each  $p$ , there exists an integer  $n$  such that, if the edges of  $K_{n,n}$  are partitioned into two sets  $A$  and  $B$ , then either  $A$  or  $B$  contains the edges of a subgraph isomorphic to  $K_{p,p}$ .

So let us assume that  $U, W \subseteq V(G)$ , where  $U$  and  $W$  are disjoint sets of  $n$  elements and there is an edge between  $u$  and  $w$  (in one or both directions) for each  $(u, w) \in U \times W$ . Let  $A$  be the set of pairs  $(u, w) \in U \times W$  such that the edge is from  $u$  to  $w$ , and let  $B = (U \times W) \setminus A$ . Then there exist sets  $U' \subseteq U$  and  $W' \subseteq W$ , with cardinality  $p$ , such that  $U' \times W' \subseteq A$  or  $W' \times U' \subseteq B$ . In either case, we get a subgraph of  $G$  isomorphic to  $\vec{K}_{p,p}$ .

Note that a quick and direct proof can be given with  $n = p2^{2p}$ , but we do not know the minimal  $n$  yielding the result.  $\square$

**Remark 6.7.** The statement relative to bounded tree-width sets of graphs in Proposition 6.5 is also a consequence (in the case of finite graphs) of Lapoire's result [31], which states that, in a graph of tree-width at most  $k$ , one can construct a width- $k$  tree-decomposition by monadic second-order (MSO) formulas. This can be used to show that every HR-recognizable set of graphs of bounded tree-width is definable in Counting Monadic Second-order (CMSO) logic, using edge set quantifications. Courcelle showed [11] that, for finite graphs of bounded tree-width, edge set quantifications can be replaced by vertex set quantifications. The considered set is therefore definable in CMSO logic with vertex set quantifications only, and hence is VR-recognizable by another of Courcelle's results [9].

**Remark 6.8.** It is proved in [8] that every set of square grids is HR-recognizable. It follows from Theorem 6.1 that every such set is also VR-recognizable. Hence, there are uncountably many VR-recognizable sets of graphs, so we cannot hope for an automata-theoretic or

a logical characterization of VR-recognizability—in contrast with the situation prevailing for words, trees and some special classes of graphs, see [43,32,33,24,29,30].

### 6.3. HR-recognizable sets which are not VR-recognizable

The aim of this short section is to establish the existence of HR-recognizable sets which are not VR-recognizable. We first establish a lemma.

**Lemma 6.9.** *Every set of cliques (of the form  $K_n$ ,  $n \geq 1$ ) is HR-recognizable.*

**Proof.** Let  $L$  be a set of undirected cliques (recall that an undirected graph is a graph where the edge relation is symmetric). We provide a locally finite  $\mathcal{CS}$ -congruence on  $\mathcal{GS}_{\text{sep}}$  which saturates  $L$  (see Section 5.2.3). By Proposition 5.7, this establishes that  $L$  is HR-recognizable.

For each finite set  $C$  of source labels, let  $G^i(C)$  be the set of graphs in  $\mathcal{GS}_{\text{sep}}(C)$  having at least one internal vertex (i.e., a vertex which is not a source), and let  $G^s(C)$  be the set of graphs in  $\mathcal{GS}_{\text{sep}}(C)$ , in which every vertex is a source. In particular,  $G^s(C)$  is finite.

Let  $\equiv$  be the following equivalence relation on  $\mathcal{GS}_{\text{sep}}$ . We use the operation  $\square_C$ , as in Section 5.2.3. If  $G, G' \in \mathcal{GS}_{\text{sep}}(C)$ , we let  $G \equiv G'$  if and only if either  $G = G'$ , or  $G, G' \in G^i(C)$  and for every  $H \in G^s(C)$ ,  $G \square_C H \in L$  iff  $G' \square_C H \in L$ .

Note that for each  $C$ , there are only finitely many  $\equiv$ -classes in  $\mathcal{GS}_{\text{sep}}(C)$ ,—namely at most  $p + 2^p$ , where  $p$  is the cardinality of  $G^s(C)$ .

Moreover,  $\equiv$  saturates  $L$ . Indeed, suppose that  $G, G' \in \mathcal{GS}_{\text{sep}}(C)$ ,  $G \equiv G'$  and  $G \in L$ . Let  $H$  be the graph in  $\mathcal{GS}_{\text{sep}}(C)$  consisting of distinct  $c$ -sources ( $c \in C$ ) and no edges. Then we have  $G = G \square_C H$  and  $G' = G' \square_C H$ . It follows from the definition of  $\equiv$  that  $G' \in L$ .

Finally, we check that  $\equiv$  is a  $\mathcal{CS}$ -congruence. Let  $G, G', H, H' \in \mathcal{GS}_{\text{sep}}(C)$ , with  $G \equiv G'$  and  $H \equiv H'$ : we want to show that  $G \square_C H \equiv G' \square_C H'$ . We observe that if both  $G$  and  $H$  have internal vertices, then  $G \square_C H$  is not a clique (by definition of operation  $\square_C$ ), and hence cannot be in  $L$ . The rest of the proof is a straightforward verification.  $\square$

We can now prove the following.

**Proposition 6.10.** *There is an HR-recognizable set of graphs which is not VR-recognizable.*

**Proof.** Let  $A$  be a set of integers which is not recognizable in  $(\mathbb{N}, \text{succ}, 0)$ , for instance the set of prime numbers, and let  $L$  be the set of cliques  $K_n$  for  $n \in A$ . Then  $L$  is HR-recognizable by Lemma 6.9.

We now consider a set of VR-terms describing  $L$  and using exactly 2 port labels,  $p$  and  $q$ . Recall that  $p$  denotes the VR-constant of type  $\{p\}$ , that is, the graph with a single vertex that is a  $p$ -port and no edges. The constant  $q$  is defined similarly. Now let  $k_1 = p$ , and  $k_{n+1} = \text{ren}_{q \rightarrow p} \text{add}_{p,q} \text{add}_{q,p}(k_n \oplus q)$ . It is not difficult to verify that  $k_n$  denotes the clique  $K_n$  where all the vertices are  $p$ -ports,  $K_n$  itself is denoted by the term  $\text{mdf}_{\emptyset} k_n$ , and the set  $K$  of all VR-terms of the form  $k_n$  is recognizable (as a set of terms, or trees). If  $L$  is VR-recognizable, then the set of VR-terms in  $K$  that denote graphs in  $L$  is recognizable. This set consists of all the terms of the form  $\text{mdf}_{\emptyset} k_n$  with  $n \in A$ , and it can be shown by standard methods that it is not recognizable. It follows that  $L$  is not VR-recognizable.  $\square$

#### 6.4. Sparse graphs and monadic second-order logic

Since graphs are relational structures, logical formulas can be used to specify sets of graphs. Monadic second-order logic is especially interesting because

*every monadic second-order definable set of finite graphs is VR-recognizable [9,15].*

There is actually a version of monadic second-order logic allowing quantifications on edges and sets of edges (one replaces the graph under consideration by its incidence graph; we omit details). We say that a set is  $MS_2$ -definable if it is definable by a monadic second-order formula with edge and edge set quantifications, and that we use the phrase  $MS_1$ -definable to refer to the first notion. It is immediately verified (from the definition) that

*Every  $MS_1$ -definable set is  $MS_2$ -definable.*

The two following statements are more difficult.

*Every  $MS_2$ -definable set of simple graphs is HR-recognizable [8].*

*If a set of simple graphs is uniformly  $k$ -sparse for some  $k$  and  $MS_2$ -definable, then it is  $MS_1$ -definable [16].*

This is somewhat analogous to the situation of Theorem 6.1 (see Proposition 6.5). However the combinatorial conditions are different: if a set of graphs is uniformly  $k$ -sparse for some  $k$ , it is without  $K_{t,t}$  for some  $t$ , but the converse does not hold. It is proved in the book by Bollobas [5] that, for each  $t \geq 2$ , there is a number  $a$  such that for each  $n$ , there is a graph with  $n$  vertices and  $an^b$  edges that does not contain  $K_{t,t}$ , where  $b = 2t/(t+1)$ . For these graphs, the number of edges is not linearly bounded in terms of the number of vertices, so they are not uniformly  $k$ -sparse for any  $k$ .

It is not clear how to extend Courcelle's proof in [16], to use the condition *without  $K_{t,t}$*  instead of *uniformly  $k$ -sparse*.

### 7. Simple graphs vs multi-graphs

The formal setting of relational structures is very convenient to deal with simple graphs, as we have seen already. It can also be used to formalize multi-graphs (i.e., graphs with multiple edges), if we consider two-sorted relational structures.

Formally, a *multi-graph with sources* in  $C$  is a structure of the form  $G = \langle V, E, \text{inc}, (c_G)_{c \in C} \rangle$  where  $V$  is the set of vertices,  $E$  is the set of edges, each  $c_G$  is an element of  $V$ , and  $\text{inc}$  is a ternary relation of type  $E \times V \times V$ . We interpret the relation  $\text{inc}(e, x, y)$  to mean that  $e$  is an edge from vertex  $x$  to vertex  $y$ . We denote by  $\mathcal{GS}_m(C)$  the set of multi-graphs with sources in  $C$ . As in the study of  $\mathcal{StS}$  or  $\mathcal{GS}$ , we assume that the finite sets of source labels  $C$  are taken in a fixed countable set. We let  $\mathcal{GS}_m$  be the union of the  $\mathcal{GS}_m(C)$  for all finite sets  $C$  of source labels.

Graphs and hypergraphs with multiple edges and hyperedges are often used, see the volume edited by Rozenberg [41]. In this context, it is in fact frequent to consider operations on multi-graphs that are very similar to the HR-operations on  $\mathcal{GS}$ . More precisely, the operations of disjoint union, source renaming, source forgetting and source fusion can be defined naturally on multigraphs with sources: thus  $\mathcal{GS}_m$  can be seen naturally as an HR-algebra.

It is clear that each simple graph in  $\mathcal{GS}(C)$  can be considered as an element in  $\mathcal{GS}_m(C)$ . It is important to note however that the HR-operations on  $\mathcal{GS}_m$ , when applied to such simple graphs, do not necessarily yield the same result as in  $\mathcal{GS}$ . For instance, let  $a, b$  be distinct elements of  $C$ , and let  $G \in \mathcal{GS}(C)$  be a simple graph. The action of fusing the  $a$ -source and the  $b$ -source of  $G$  may now result in multiple edges: if there were arrows in both directions between  $a_G$  and  $b_G$ , or if there were arrows to (resp. from) a vertex of  $G$  from (resp. to) both  $a_G$  and  $b_G$ . In contrast, the same operation in  $\mathcal{GS}(C)$  yields  $\text{fus}_{a,b}(G)$ , an element of  $\mathcal{GS}(C)$  by definition. To avoid confusion, we will denote by  $\text{mfus}_{a,b}$  this operation when used in  $\mathcal{GS}_m$ .

Fortunately, we do not have this sort of problem with the other operations: applying the operations of disjoint union, source renaming or source forgetting to simple graphs considered as elements of  $\mathcal{GS}_m$  yields the same result as applying the same operations within the algebra  $\mathcal{GS}$ .

We let  $\text{HR}_m$  be the signature on  $\mathcal{GS}_m$  consisting of the operations of the form  $\oplus$ ,  $\text{srcfg}_a$ ,  $\text{srcren}_{a \rightarrow b}$  and  $\text{mfus}_{a,b}$ . Thus,  $\mathcal{GS}_m$  is an  $\text{HR}_m$ -algebra. We observe that, as a signature (that is, as a set of symbols denoting operations),  $\text{HR}_m$  is in natural bijection with  $\text{HR}$ . So we do not really need to introduce the new notation  $\text{HR}_m$ , and we could very well say that  $\mathcal{GS}_m$  is an  $\text{HR}$ -algebra. We simply hope, by introducing this notation, to clarify our comparative study of recognizable subsets in the algebras  $\mathcal{GS}$  and  $\mathcal{GS}_m$ . This distinction will be useful in the proofs of Theorems 7.3 and 7.4.

To summarize and amplify the above remarks, let us introduce the following notation. We denote by  $\iota: \mathcal{GS} \rightarrow \mathcal{GS}_m$  the natural injection. For each multi-graph  $G$ , we denote by  $u(G)$  the simple graph obtained from  $G$  by fusing multiple edges (with identical origin and end): that is,  $u$  is a mapping from  $\mathcal{GS}_m$  onto  $\mathcal{GS}$ . Elementary properties of  $\iota$  and  $u$  are listed in the next proposition.

**Proposition 7.1.** *The mapping  $u: \mathcal{GS}_m \rightarrow \mathcal{GS}$  is a homomorphism of  $\text{HR}$ -algebras. The mapping  $\iota: \mathcal{GS} \rightarrow \mathcal{GS}_m$  is not a homomorphism, but it commutes with the operations of the form  $\oplus$ ,  $\text{srcfg}_a$  and  $\text{srcren}_{a \rightarrow b}$ .*

*$\iota$  does not commute with the operations of the form  $\text{fus}_{a,b}$ , but if  $G \in \mathcal{GS}$ , then  $\iota(\text{fus}_{a,b}(G)) = \iota(u(\text{mfus}_{a,b}(\iota(G))))$ .*

*Finally, if  $G \in \mathcal{GS}$ , then  $\iota(G) = u^{-1}(G) \cap \iota(\mathcal{GS})$  and  $u(\iota(G)) = G$ .*

We now prove the following theorems, which describe the interaction between  $\text{HR}_m$ -recognizability of sets of multi-graphs and  $\text{HR}$ -recognizability of sets of simple graphs.

**Theorem 7.2.** *The set of simple graphs is  $\text{HR}_m$ -recognizable. More precisely, for each finite set of source labels  $C$ ,  $\iota(\mathcal{GS}(C))$  is  $\text{HR}_m$ -recognizable.*

**Theorem 7.3.** *Let  $C$  be a finite set of source labels and let  $L \subseteq \mathcal{GS}(C)$ . Then  $L$  is  $\text{HR}$ -recognizable if and only if  $\iota(L)$  is  $\text{HR}_m$ -recognizable.*

**Theorem 7.4.** *Let  $C$  be a finite set of source labels and let  $L \subseteq \mathcal{GS}_m(C)$ . If  $L$  is  $\text{HR}_m$ -recognizable, then  $u(L)$  is  $\text{HR}$ -recognizable.*

### 7.1. Proof of Theorem 7.2

We first introduce the notion of the *type* of a multi-graph: as for the elements of  $StS$ , if  $G \in \mathcal{GS}_m(C)$ , we let  $\zeta(G)$  be the restriction of  $G$  to its  $C$ -sources and to the edges between them. We also denote by  $\zeta$  the relation on  $\mathcal{GS}_m$  induced by this type mapping: two multi-graphs  $G, H \in \mathcal{GS}_m(C)$  are  $\zeta$ -equivalent if  $\zeta(G) = \zeta(H)$ .

**Lemma 7.5.** *The type relation  $\zeta$  is an  $HR_m$ -congruence on  $\mathcal{GS}_m$ . Moreover, for each finite set of source labels  $C$ , the elements of  $\iota(\mathcal{GS}(C))$  can be found in only a finite number of  $\zeta$ -classes.*

**Proof.** The result follows from the following, easily verifiable identities, where the multi-graphs  $G, H$  are assumed to have the appropriate sets of sources.

$$\begin{aligned}\zeta(G \oplus H) &= \zeta(G) \oplus \zeta(H), \\ \zeta(\text{srcren}_{a \rightarrow b}(G)) &= \text{srcren}_{a \rightarrow b}(\zeta(G)), \\ \zeta(\text{mfus}_{a,b}(G)) &= \text{mfus}_{a,b}(\zeta(G)), \\ \zeta(\text{srcfg}_a(G)) &= \zeta(\text{srcfg}_a(\zeta(G))).\end{aligned}$$

The finiteness of the number of  $\zeta$ -classes containing elements of  $\iota(\mathcal{GS}(C))$  follows from the fact that there are only finitely many source-only simple graphs with sources in  $C$ .  $\square$

We also introduce the following finite invariant for a simple graph  $G \in \mathcal{GS}(C)$ . We define  $\eta(G)$  to be the set of all pairs  $\{a, b\}$  of elements of  $C$  such that  $a \neq b$ ,  $a_G \neq b_G$  and there exists a vertex  $x$  of  $G$  with either edges from  $x$  to both  $a_G$  and  $b_G$ , or edges to  $x$  from both  $a_G$  and  $b_G$ . The set  $\eta(G)$  can be viewed as a symmetric anti-reflexive relation on  $C$ .

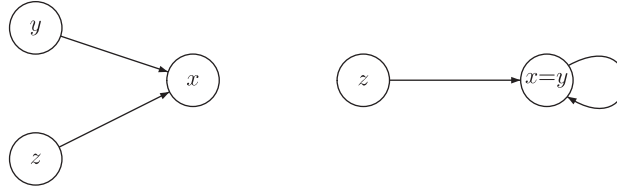
**Lemma 7.6.** *Let  $G$  be a simple graph in  $\mathcal{GS}(C)$  and let  $a \neq b$  be elements of  $C$ . Then  $\text{mfus}_{a,b}(G)$  has multiple edges if and only if  $\{a, b\} \in \eta(G)$  or  $\text{mfus}_{a,b}(\zeta(G))$  has multiple edges.*

**Proof.** We first observe that  $\text{mfus}_{a,b}(G)$  has multiple edges if and only if  $a_G \neq b_G$  and at least one of the following situations occurs: there are edges in both directions between  $a_G$  and  $b_G$ , or there is a vertex  $x$  of  $G$  with edges from (resp. to) both  $a_G$  and  $b_G$  (this includes the case where there is a loop at  $a_G$  or  $b_G$  and an edge in either direction between  $a_G$  and  $b_G$ ). That is,  $\text{mfus}_{a,b}(G)$  has multiple edges if and only if  $\{a, b\} \in \eta(G)$  or there are edges in both directions between  $a_G$  and  $b_G$ .

We also observe that  $\text{mfus}_{a,b}(\zeta(G))$  is a subgraph of  $\text{mfus}_{a,b}(G)$ , so the former is simple if the latter is. Finally, the existence of edges in both directions between  $a_G$  and  $b_G$  is sufficient to ensure that  $\text{mfus}_{a,b}(\zeta(G))$  has multiple edges.

These observations put together suffice to prove the lemma.  $\square$

We are now ready to prove Theorem 7.2. Let  $\simeq$  be the following relation, defined on each  $\mathcal{GS}_m(C)$ . We let  $G \simeq G'$  if both  $G$  and  $G'$  have multiple edges, or both  $G$  and  $G'$  are simple graphs,  $\zeta(G) = \zeta(G')$  and  $\eta(G) = \eta(G')$ .

Fig. 4. Distinct edges in  $\text{mfus}_{a,b}(G)$ .

It is immediate that  $\simeq$  is an equivalence relation, saturating  $\iota(\mathcal{GS}(C))$ . It follows from Lemma 7.5 and from the fact that  $\eta(G)$  is a subset of the finite set  $C \times C$ , that  $\simeq$  is locally finite. So we only need to show that  $\simeq$  is an  $\text{HR}_m$ -congruence.

We need to describe the interaction between the mapping  $\eta$  and the  $\text{HR}_m$ -operations. As observed in Proposition 7.1, all  $\text{HR}_m$ -operations preserve simple graphs except for the operations of the form  $\text{mfus}_{a,b}$ . Assuming that  $G, H$  are simple graphs with the appropriate sets of sources, we easily verify the following:

$$\begin{aligned}\eta(G \oplus H) &= \eta(G) \cup \eta(H), \\ \eta(\text{srcfg}_a(G)) &= \eta(G) \setminus \{\{a, b\} \mid b \in C, \{a, b\} \in \eta(G)\}, \\ \eta(\text{srcren}_{a \rightarrow b}(G)) &= \eta(G) \setminus \{\{a, c\} \mid c \in C, \{a, c\} \in \eta(G)\} \\ &\quad \cup \{\{b, c\} \mid c \in C, \{a, c\} \in \eta(G)\}.\end{aligned}$$

Moreover, if  $a_G \neq b_G$  and  $\text{mfus}_{a,b}(G)$  is simple (if it isn't, its  $\eta$ -image is not defined), then  $\eta(\text{mfus}_{a,b}(G))$  consists of:

- (1) all pairs in  $\eta(G)$ ,
- (2) all pairs  $\{c, d\}$  such that there are edges in  $\zeta(G)$  from  $a$  to  $c$  and from  $b$  to  $d$ , or from  $c$  to  $a$  and from  $d$  to  $b$ ,
- (3) all pairs  $\{a, c\}$  (resp.  $\{b, c\}$ ) such that  $\{b, c\} \in \eta(G)$  (resp.  $\{a, c\} \in \eta(G)$ ),
- (4) all pairs  $\{a, c\}$  and  $\{b, c\}$  such that there are edges in  $\zeta(G)$  between  $a$  and  $b$  (in either direction) and between  $a$  or  $b$  and  $c$  (in any direction).

Let us justify this statement: it is easy to see that all these pairs belong to  $\eta(\text{mfus}_{a,b}(G))$ . In particular,  $\eta(G) \subseteq \eta(\text{mfus}_{a,b}(G))$  since, as  $\text{mfus}_{a,b}(G)$  is assumed to be simple, there is no  $\{c, d\} \in \eta(G)$  such that  $a_G = c_G$  and  $b_G = d_G$ .

Conversely, let us consider distinct edges in  $G' = \text{mfus}_{a,b}(G)$ , from  $y$  to  $x$  and from  $z$  to  $x$ , as in Fig. 4 (note that  $x$  and  $y$  may be equal), such that  $y = e_{G'}$  and  $z = f_{G'}$  for  $e, f \in C$ . If neither  $x$ , nor  $y$  nor  $z$  is the  $a$ - and  $b$ -source in  $G'$ , then we are in case (1), i.e.,  $\{e, f\} \in \eta(G)$ . If  $x$  is the  $a$ - and  $b$ -source in  $G'$  but neither  $y$  nor  $z$  is, then  $\{e, f\}$  satisfies case (1) or (2). If  $y$  is the  $a$ - and  $b$ -source in  $G'$  but neither  $x$  nor  $z$  is, then  $\{e, f\}$  satisfies case (3). The same holds by symmetry if  $z$  is the only one of these three vertices to be the  $a$ - and  $b$ -source in  $G'$ . Finally if  $x = y$  (resp.  $x = z$ ) and is the  $a$ - and  $b$ -source, in  $G'$  then there is an edge between the  $a$ - and the  $b$ -source in  $G$  and  $\{e, f\}$  satisfies case (4). The case of edges from  $x$  to  $y$  and to  $z$  is symmetrical.

In particular,  $\eta(G \oplus H)$ ,  $\eta(\text{srcfg}_a(G))$ ,  $\eta(\text{srcren}_{a \rightarrow b}(G))$  and  $\eta(\text{mfus}_{a,b}(G))$  are entirely determined by  $\eta(G)$ ,  $\zeta(G)$  and  $\eta(H)$ .

Let us now consider  $G, G', H, H'$  in  $\mathcal{GS}_m$  (with the appropriate sets of sources) such that  $G \simeq G'$  and  $H \simeq H'$ . If  $G$  is not simple, then neither are  $G', G \oplus H, \text{srcfg}_a(G), \text{srcren}_{a \rightarrow b}(G)$  and  $\text{mfus}_{a,b}(G)$ . In particular, we have  $G \oplus H \simeq G' \oplus H', \text{srcfg}_a(G) \simeq \text{srcfg}_a(G'), \text{srcren}_{a \rightarrow b}(G) \simeq \text{srcren}_{a \rightarrow b}(G')$  and  $\text{mfus}_{a,b}(G) \simeq \text{mfus}_{a,b}(G')$ .

Assume now that  $G$  and  $H$  are simple. Then so are  $G \oplus H, \text{srcfg}_a(G)$  and  $\text{srcren}_{a \rightarrow b}(G)$ , and we have seen that their  $\eta$ -images are determined by  $\eta(G)$  and  $\eta(H)$ . Since  $\zeta$  is an  $\text{HR}_m$ -congruence (Lemma 7.5), it follows that  $\simeq$  is preserved by the operations  $\oplus, \text{srcfg}_a$  and  $\text{srcren}_{a \rightarrow b}$ .

By Lemma 7.6, whether  $\text{mfus}_{a,b}(G)$  is simple, is determined by  $\zeta(G)$  and  $\eta(G)$ , and hence  $\text{mfus}_{a,b}(G)$  and  $\text{mfus}_{a,b}(G')$  are both non-simple (and then  $\simeq$ -equivalent) or both simple. In the latter case, their  $\eta$ -images are equal since they are both determined by  $\eta(G) = \eta(G')$  and  $\zeta(G) = \zeta(G')$ . Thus  $\simeq$  is preserved by the operation  $\text{mfus}_{a,b}$ . This concludes the proof of Theorem 7.2.

### 7.2. Proof of Theorem 7.3

Recall that we want to show that for each  $L \in \mathcal{GS}(C)$ ,  $L$  is  $\text{HR}$ -recognizable if and only if  $\iota(L)$  is  $\text{HR}_m$ -recognizable.

One direction is quickly established: we know from Proposition 7.1 that  $\iota(L) = u^{-1}(L) \cap \iota(\mathcal{GS}(C))$ . If  $L$  is  $\text{HR}$ -recognizable, then  $u^{-1}(L)$  is  $\text{HR}_m$ -recognizable since  $u$  is a homomorphism. In view of Theorem 7.2, it follows that  $\iota(L)$  is  $\text{HR}_m$ -recognizable as well.

Conversely, let us assume that  $\iota(L)$  is  $\text{HR}_m$ -recognizable and let  $\equiv$  be a locally finite  $\text{HR}_m$ -congruence on  $\mathcal{GS}_m$  saturating  $\iota(L)$ . We want to define a locally finite  $\text{HR}$ -congruence  $\sim$  on  $\mathcal{GS}$  saturating  $L$ .

For each symmetric anti-reflexive relation  $A$  on a finite set of source labels  $D$  and for each graph  $G \in \mathcal{GS}(D)$ , let  $\text{del}_A(G) \in \mathcal{GS}(D)$  be the graph obtained from  $G$  by deleting the edges between the  $a$ -source and the  $b$ -source for each pair  $\{a, b\}$  in  $D$ . Let also  $\text{fus}_A$  be the composition of the operations  $\text{fus}_{a,b}$  for all  $\{a, b\} \in A$ , in any order.

For  $G, G' \in \mathcal{GS}(D)$ , we let  $G \sim G'$  if  $\iota(G) \equiv \iota(G'), \zeta(G) = \zeta(G')$  and, for each symmetric anti-reflexive relation  $A$  on  $D$ ,

$$\iota \text{fus}_A \text{del}_A(G) \equiv \iota \text{fus}_A \text{del}_A(G').$$

The relation  $\sim$  is clearly an equivalence relation, and it is locally finite since  $\equiv$  and  $\zeta$  are. Moreover, it saturates  $L$  since  $G \in L$  if and only if  $\iota(G) \in \iota(L)$ , and  $\equiv$  saturates  $\iota(L)$ . The rest of the proof consists in showing that  $\sim$  is an  $\text{HR}$ -congruence.

*The source renaming operation.* Let  $G \sim G'$  in  $\mathcal{GS}(D)$ . Then  $\iota(G) \equiv \iota(G')$ . Since  $\equiv$  is a congruence and in view of Proposition 7.1,  $\iota(\text{srcren}_{a \rightarrow b}(G)) = \text{srcren}_{a \rightarrow b}(\iota(G)) \equiv \text{srcren}_{a \rightarrow b}(\iota(G')) = \iota(\text{srcren}_{a \rightarrow b}(G'))$ . It also follows from Lemma 3.9 that  $\zeta(\text{srcren}_{a \rightarrow b}(G)) = \zeta(\text{srcren}_{a \rightarrow b}(G'))$ .

Let us now consider a symmetric anti-reflexive relation  $A$  on the set of source labels of  $\text{srcren}_{a \rightarrow b}(G)$ . It is easily verified that

$$\text{del}_A \text{srcren}_{a \rightarrow b} = \text{srcren}_{a \rightarrow b} \text{del}_B,$$

where  $B = \{\{c, d\} \in A \mid \{c, d\} \cap \{a, b\} = \emptyset\} \cup \{\{a, d\} \mid \{b, d\} \in A\}$ . We also note that if  $c, d \in C \setminus \{a, b\}$ , then  $\text{fus}_{c,d}$  and  $\text{srcren}_{a \rightarrow b}$  commute. Moreover  $\text{fus}_{b,d} \text{srcren}_{a \rightarrow b} = \text{srcren}_{a \rightarrow b} \text{fus}_{a,d}$  and  $\text{fus}_{c,b} \text{srcren}_{a \rightarrow b} = \text{srcren}_{a \rightarrow b} \text{fus}_{c,a}$ . Thus  $\text{fus}_A \text{srcren}_{a \rightarrow b} = \text{srcren}_{a \rightarrow b} \text{fus}_B$ .

Now, using the fact that  $\iota$  commutes with  $\text{srcren}_{a \rightarrow b}$  we have

$$\begin{aligned} \iota \text{fus}_A \text{del}_A \text{srcren}_{a \rightarrow b}(G) &= \iota \text{fus}_A \text{srcren}_{a \rightarrow b} \text{del}_B(G) \\ &= \iota \text{srcren}_{a \rightarrow b} \text{fus}_B \text{del}_B(G) \\ &= \text{srcren}_{a \rightarrow b} \iota \text{fus}_B \text{del}_B(G). \end{aligned}$$

Since  $\equiv$  is an  $\text{HR}_m$ -congruence, it follows that

$$\iota \text{fus}_A \text{del}_A \text{srcren}_{a \rightarrow b}(G) \equiv \iota \text{fus}_A \text{del}_A \text{srcren}_{a \rightarrow b}(G')$$

and, finally, that  $\text{srcren}_{a \rightarrow b}(G) \sim \text{srcren}_{a \rightarrow b}(G')$ .

*The source forgetting operation.* The proof is the same as for the source renaming operation, with this simplifying circumstance that  $\text{del}_A \text{srcfg}_a = \text{srcfg}_a \text{del}_A$  and  $\text{fus}_A \text{srcfg}_a = \text{srcfg}_a \text{fus}_A$  (since  $a$  is not a source label of  $\text{srcfg}_a(G)$ , and hence does not occur in  $A$ ).

*The source fusion operation.* Let  $G \sim G'$  in  $\mathcal{GS}(D)$ . Here it is not immediate that  $\iota(\text{fus}_{a,b}(G)) \equiv \iota(\text{fus}_{a,b}(G'))$ . However, if we let  $A = \{\{a, b\}\}$ , we know that

$$\iota \text{fus}_A \text{del}_A(G) \equiv \iota \text{fus}_A \text{del}_A(G').$$

We note that  $\text{fus}_A \text{del}_A(G)$  is equal to  $\text{fus}_{a,b}(G)$  if  $G$  has no edge between its  $a$ - or  $b$ -source, or if it has a loop at either. Otherwise,  $\text{fus}_{a,b}(G)$  is equal to  $\text{fus}_A \text{del}_A(G)$  with a loop added to its  $a$ -source, that is:

$$\text{fus}_{a,b}(G) = \text{srcfg}_\alpha \text{srcfg}_\beta \text{fus}_{a,\alpha} \text{fus}_{b,\beta} (\text{fus}_A \text{del}_A(G) \oplus E), \quad (*)$$

where  $\alpha$  and  $\beta$  are source labels not in  $D$  and  $E$ , is the graph in  $\mathcal{GS}(\{\alpha, \beta\})$  with 2 vertices and a single edge from its  $\alpha$ -source to its  $\beta$ -source.

Observe also that the existence of loops at, or edges between the  $a$ - and  $b$ -source of  $G$  is a condition that depends only on  $\zeta(G)$ , so it will be satisfied by both  $G$  and  $G'$  or by neither.

In the first case, where  $\text{fus}_A \text{del}_A(G) = \text{fus}_{a,b}(G)$ , we find immediately that  $\iota(\text{fus}_{a,b}(G)) \equiv \iota(\text{fus}_{a,b}(G'))$ . In the second case, the same  $\equiv$ -equivalence is derived from Proposition 7.1 and Eq. (\*) above.

By Lemma 3.9,  $\zeta$ -equivalence is preserved by the operation  $\text{fus}_{a,b}$ .

Now let  $A$  be a symmetric anti-reflexive relation on  $D$ : we consider the graph  $\iota \text{fus}_A \text{del}_A \text{fus}_{a,b}(G)$ . Our first observation is that  $\text{del}_A \text{fus}_{a,b} = \text{fus}_{a,b} \text{del}_B$  where

$$B = A \cup \{\{a, c\} \mid \{b, c\} \in A\} \cup \{\{b, c\} \mid \{a, c\} \in A\}.$$

Next, we observe that  $\text{fus}_A \text{fus}_{a,b} = \text{fus}_{a,b} \text{fus}_B$ . Thus we have

$$\iota \text{fus}_A \text{del}_A \text{fus}_{a,b}(G) = \iota \text{fus}_A \text{fus}_{a,b} \text{del}_B = \iota \text{fus}_{a,b} \text{del}_B \text{fus}_B(G),$$

and hence  $\imath \text{fus}_A \text{del}_A \text{fus}_{a,b}(G) \equiv \imath \text{fus}_A \text{del}_A \text{fus}_{a,b}(G')$ . It follows that  $\text{fus}_{a,b}(G) \sim \text{fus}_{a,b}(G')$ .

*The disjoint union operation.* Let  $G \sim G'$  in  $\mathcal{GS}(C)$  and  $H \sim H'$  in  $\mathcal{GS}(D)$  (where  $C$  and  $D$  are disjoint). Since  $\imath$  and  $\zeta$  preserve  $\oplus$ , we have  $\imath(G \oplus H) \equiv \imath(G' \oplus H')$  and  $\zeta(G \oplus H) = \zeta(G' \oplus H')$ .

Now let  $A$  be a symmetric anti-reflexive relation on  $C \cup D$ . Let  $Q$  (resp.  $R$ ) be the restriction of  $A$  to  $C$  (resp.  $D$ ) and let  $P = A \cap ((C \times D) \cup (D \times C))$ . It is easily verified that

$$\begin{aligned} \text{del}_A(G \oplus H) &= \text{del}_Q(G) \oplus \text{del}_R(H) \\ \text{fus}_A \text{del}_A(G \oplus H) &= \text{fus}_P(\text{fus}_Q \text{del}_Q(G) \oplus \text{fus}_R \text{del}_R(H)). \end{aligned}$$

It now follows from Proposition 7.1 that

$$\begin{aligned} \imath \text{fus}_A \text{del}_A(G \oplus H) &= \imath \text{fus}_P(\text{fus}_Q \text{del}_Q(G) \oplus \text{fus}_R \text{del}_R(H)) \\ &= \imath \text{umfus}_P \imath(\text{fus}_Q \text{del}_Q(G) \oplus \text{fus}_R \text{del}_R(H)) \\ &= \imath \text{umfus}_P(\imath \text{fus}_Q \text{del}_Q(G) \oplus \imath \text{fus}_R \text{del}_R(H)). \end{aligned}$$

Thus  $\imath \text{fus}_A \text{del}_A(G \oplus H) \equiv \imath \text{fus}_A \text{del}_A(G' \oplus H')$ , and hence  $G \oplus H \sim G' \oplus H'$ .

This concludes the proof of Theorem 7.3.

### 7.3. Proof of Theorem 7.4

Let  $L \in \mathcal{GS}_m(C)$  be  $\text{HR}_m$ -recognizable, and let  $\equiv$  be a locally finite  $\text{HR}_m$ -congruence saturating  $L$ . We want to show that  $u(L)$  (a subset of  $\mathcal{GS}(C)$ ) is  $\text{HR}$ -recognizable.

Let  $G, G' \in \mathcal{GS}(D)$ . We let  $G \sim G'$  if, for each  $H \in u^{-1}(G)$ , there exists  $H' \in u^{-1}(G')$  such that  $H \equiv H'$ , and symmetrically, for each  $H' \in u^{-1}(G')$ , there exists  $H \in u^{-1}(G)$  such that  $H \equiv H'$ .

The relation  $\sim$  is easily seen to be a locally finite equivalence relation on  $\mathcal{GS}$ , saturating  $u(L)$ . There remains to see that  $\sim$  is an  $\text{HR}$ -congruence.

We first establish the following lemma.

**Lemma 7.7.** *Let  $G \in \mathcal{GS}_m$  and let  $H, K \in \mathcal{GS}$ .*

- $u(G) = H \oplus K$  if and only if there exist multi-graphs  $H', K'$  such that  $G = H' \oplus K'$ ,  $u(H') = H$  and  $u(K') = K$ .
- $u(G) = \text{srcfg}_a(H)$  if and only if there exists a multi-graph  $H'$  such that  $G = \text{srcfg}_a(H')$  and  $u(H') = H$ .
- $u(G) = \text{srcren}_{a \rightarrow b}(H)$  if and only if there exists a multi-graph  $H'$  such that  $G = \text{srcren}_{a \rightarrow b}(H')$  and  $u(H') = H$ .
- $u(G) = \text{fus}_{a,b}(H)$  if and only if there exists a multi-graph  $H'$  such that  $G = \text{mfus}_{a,b}(H')$  and  $u(H') = H$ .

**Proof.** Recall that  $G$  and  $u(G)$  have the same set of vertices, and each edge  $e$  of  $u(G)$  arises from the identification  $n(e) \geq 1$  edges of  $G$  between the same vertices.

If  $u(G) = H \oplus K$ , each edge of  $u(G)$  is in exactly one of  $H$  and  $K$ . Let  $H'$  (resp.  $K'$ ) be the graph obtained from  $H$  (resp.  $K$ ) by replacing each edge  $e$  by  $n(e)$  parallel edges. Then  $G = H' \oplus K'$ ,  $u(H') = H$  and  $u(K') = K$ , as required.

The proof of the statements relative to the operations  $\text{srcfg}_a$  and  $\text{srcren}_{a \rightarrow b}$  is done in the same fashion.

Let us finally consider the case where  $u(G) = \text{fus}_{a,b}(H)$ . If  $a_H = b_H$ , that is,  $H = u(G)$ , then  $G = \text{mfus}_{a,b}(G)$  and we can let  $H' = G$ .

If  $a_H \neq b_H$ , we let  $H'$  be obtained from  $H$  as follows: for each vertex  $x$ , each edge  $e$  from  $x$  to  $y$  ( $y \neq a, b$ ) is replaced by  $n(e)$  parallel edges, and the edges from  $x$  to  $a$  and  $b$  are duplicated to a total of  $n(e)$  edges.  $\square$

We can now conclude the proof of Theorem 7.4, by proving that  $\sim$  is an HR-congruence. Let  $G \sim G'$  and  $H \sim H'$ . Let  $K \in u^{-1}(G \oplus H)$ . By Lemma 7.7,  $K = L \oplus M$  for some  $L \in u^{-1}(G)$  and  $M \in u^{-1}(H)$ . Since  $G \sim G'$  and  $H \sim H'$ , there exist  $L' \in u^{-1}(G')$  and  $M' \in u^{-1}(H')$  such that  $L' \equiv L$  and  $M' \equiv M$ . Let  $K' = L' \oplus M'$ . Then  $K' = L' \oplus M' \equiv L \oplus M = K$  and  $K' \in u^{-1}(G' \oplus H')$ . By symmetry, this shows that  $G \oplus H \sim G' \oplus H'$ .

The verification that  $\sim$  is preserved by the other HR-operations proceeds along the same lines. This concludes the proof of Theorem 7.4.

## 8. Graph algebras based on graph substitutions

The class **Graph**, defined in Section 3.1, has already been discussed in terms of the signatures  $\mathcal{S}$ ,  $\text{VR}$  and  $\text{HR}$  since it is a domain in each of the three algebras  $\text{StS}$ ,  $\mathcal{GP}$  and  $\mathcal{GS}$ . In this section, we consider a different set of operations on **Graph**, arising from the theory of the modular decomposition of graphs, which makes **Graph** an algebra (one-sorted for a change!). This algebraic framework was considered by the authors, in [14,46].

We first recall the definition of the composition operation on graphs. Let  $H$  be a graph with vertex set  $[n] = \{1, \dots, n\}$  ( $n \geq 2$ ). If  $G_1, \dots, G_n$  are graphs, then the graph  $H(G_1, \dots, G_n)$  is obtained by taking the disjoint union of the graphs  $G_1, \dots, G_n$ , and by adding, for each edge  $(i, j)$  of  $H$  where  $i \neq j$ , an edge from every vertex of  $G_i$  to every vertex of  $G_j$ .

We say that a graph is *indecomposable*, or *prime*, if it cannot be written non-trivially as a composition (a composition is trivial if each of its arguments is a singleton). It is easily verified that if  $H$  and  $H'$  are isomorphic graphs, then the corresponding composition operations yield isomorphic graphs. So we fix a set  $\mathcal{F}_\infty$  of representatives of the isomorphism classes of indecomposable graphs. In particular, we may assume that every graph in  $\mathcal{F}_\infty$  has a vertex set of the form  $[n]$  for some  $n \geq 2$ . We also denote by  $\mathcal{F}_\infty$  the resulting *modular signature*, consisting of the composition operations defined by these graphs. The  $\mathcal{F}_\infty$ -algebra of graphs is denoted by  $\text{Graph}^{\mathcal{F}_\infty}$ .

It turns out that every finite graph admits a *modular decomposition*, that is, it can be expressed from the single-vertex graph using only operations from  $\mathcal{F}_\infty$ . This fact has been rediscovered a number of times in the context of graph theory and of other fields using graph-theoretic representations. We refer to [38] for a historical survey, and to [36] for a concise presentation. In other words, **Graph** is generated by the signature  $\mathcal{F}_\infty$  augmented with the constants  $\mathbf{v}^{\text{loop}}$  and  $\mathbf{v}$ , which denote a single vertex graph, respectively with and without a single loop edge.

**Remark 8.1.** The modular decomposition of a graph is unique up to certain simple (equational) rules, see for instance [46]. Moreover, the modular decomposition of a graph can be computed in linear time [35,36,21].

Our first results connect VR-recognizability and  $\mathcal{F}_\infty$ -recognizability.

**Proposition 8.2.** *Every VR-recognizable set of graphs is  $\mathcal{F}_\infty$ -recognizable.*

**Proof.** In view of Proposition 2.1 and Theorem 4.5, it suffices to show that every operation in  $\mathcal{F}_\infty$  is  $\text{VR}^+$ -derived.

For each integer  $i$ , let  $\text{mark}_i$  be the unary operation on  $\mathcal{GP}$ , of type  $\emptyset \rightarrow \{i\}$ , defined as follows: given a graph without ports, it simply marks every vertex with port label  $i$  (leaving the set of vertices and the edge relation unchanged). Note that  $\text{mark}_i$  is a qfd unary operation, and hence a  $\text{VR}^+$ -operation.

Let  $H$  be an  $n$ -ary operation, that is, a graph in  $\mathcal{F}_\infty$  with vertex set  $[n]$ , and let  $\text{edge}_H$  be its edge relation. If  $G_1, \dots, G_n$  are finite graphs, the construction of  $H\langle G_1, \dots, G_n \rangle$  can be described as follows:

- construct the disjoint union,  $\text{mark}_1(G_1) \oplus \dots \oplus \text{mark}_n(G_n)$ , an element of  $\mathcal{GP}([n])$ ;
- apply (in any order) to this disjoint union the operations  $\text{add}_{i,j}$  for all  $i, j \in [n]$  such that  $(i, j)$  is an edge of  $H$  and  $i \neq j$ ;
- forget all ports, that is, apply the operation  $\text{mdf}_\emptyset$ .

This completes the verification that the operation defined by  $H$  can be expressed as a  $\text{VR}^+$ -term, and hence the proof.  $\square$

The following result shows that the converse of Proposition 8.2 does not hold.

**Proposition 8.3.** *Every set of prime graphs is  $\mathcal{F}_\infty$ -recognizable, and there is a set of prime graphs which is not VR-recognizable.*

**Proof.** Let  $L$  be a set of prime graphs, and let  $\equiv$  be the relation on  $\text{Graph}$  defined as follows. We let  $G \equiv H$  if one of the following holds:

- neither  $G$  nor  $H$  is prime;
- $G$  and  $H$  are both  $\mathbf{1}$  (the graph with one vertex and no edge);
- $G$  and  $H$  are both not  $\mathbf{1}$ , prime and in  $L$ ;
- $G$  and  $H$  are both not  $\mathbf{1}$ , prime and not in  $L$ .

This is clearly an equivalence relation with four classes, which saturates  $L$ . Moreover,  $\equiv$  is an  $\mathcal{F}_\infty$ -congruence. Indeed, let  $K$  be a graph with  $n$  vertices; for  $i = 1, \dots, n$ , let  $G_i \equiv H_i$  for each  $i$ . If for some  $i$ ,  $G_i \neq \mathbf{1}$ , then  $H_i \neq \mathbf{1}$ , and neither  $K\langle G_1, \dots, G_n \rangle$  nor  $K\langle H_1, \dots, H_n \rangle$  is prime: therefore they are equivalent. Otherwise,  $G_i = H_i = \mathbf{1}$  for each  $i$ ,  $K\langle G_1, \dots, G_n \rangle$  and  $K\langle H_1, \dots, H_n \rangle$  are both equal to  $K$ , and hence they are equivalent. This concludes the proof that every set of prime graphs is  $\mathcal{F}_\infty$ -recognizable.

Before we exhibit a set of prime graphs which is not VR-recognizable, we define inductively a sequence of VR-terms written with three port labels  $a, b, c$ . We let

$$t_0 = \text{add}_{a,b}(a \oplus b), \quad t_{n+1} = \text{ren}_{c \rightarrow b}(\text{ren}_{b \rightarrow a}(\text{add}_{b,c}(t_n \oplus c))).$$

The term  $\text{mdf}_{\emptyset}(t_n)$  (forgetting all port labels in  $t_n$ ) denotes the string graph  $P_{n+2}$ , with  $n+2$  vertices, say  $1, \dots, n+2$  and edges from  $i$  to  $i+1$  for each  $1 \leq i \leq n+1$ . Each of these graphs is prime.

Now let  $A$  be a set of positive integers that is not recognizable in  $\langle \mathbb{N}, \text{succ}, 0 \rangle$  and let  $L$  be the set of all terms  $P_n$  with  $n \in A$ . From the above discussion, we know that  $L$  is  $\mathcal{F}_{\infty}$ -recognizable. If  $L$  was VR-recognizable, standard arguments would show that the set of VR-terms  $t_n$  ( $n \in A$ ) would be recognizable as well, and it would follow that  $A$  is recognizable, contradicting its choice.  $\square$

Now let  $\mathcal{F}$  be a finite subsignature of the modular signature  $\mathcal{F}_{\infty}$ . A graph which can be constructed from one-vertex graphs using only operations from  $\mathcal{F}$  is called an  $\mathcal{F}$ -graph. The next result deals with sets of  $\mathcal{F}$ -graphs. This finiteness condition (the elements of  $L$  are built by repeated composition of a finite number of graph-based operations) is non-trivial. In fact, for many natural classes of graphs such as rectangular grids, it is not satisfied: since grids are indecomposable, a set of graphs containing infinitely many grids cannot satisfy our finiteness condition. But that condition is satisfied by other classical classes (e.g. cographs, series-parallel posets), see [14,46].

Using results of Courcelle [14], we can show the following result, which yields in particular a weak converse of Proposition 8.2.

**Theorem 8.4.** *Let  $\mathcal{F}$  be a finite subsignature of  $\mathcal{F}_{\infty}$  and let  $L$  be a set of  $\mathcal{F}$ -graphs. The following properties are equivalent:*

1.  $L$  is  $\mathcal{S}$ -recognizable;
2.  $L$  is VR-recognizable.
3.  $L$  is  $\mathcal{F}_{\infty}$ -recognizable.
4.  $L$  is  $\mathcal{F}$ -recognizable.

**Proof.** The equivalence of (1) and (2) can be found in Theorem 4.5. Proposition 8.2 shows that (2) implies (3). And (3) implies (4) as an immediate consequence of Proposition 2.1 since  $\mathcal{F}$  is a subsignature of  $\mathcal{F}_{\infty}$ . The fact that (4) implies (1) is a consequence of two results of Courcelle: [14, Theorem 4.1], which states that if a set of  $\mathcal{F}$ -graphs is  $\mathcal{F}$ -recognizable, then it is definable in a certain extension of MS-logic; and [14, Theorem 6.11], which states that all sets definable in this logical language are  $\mathcal{S}$ -recognizable.  $\square$

**Remark 8.5.** Theorem 8.4 states that for sets of graphs with only finitely many prime subgraphs, all four notions of recognizability are equivalent. Presented in this fashion, the statement is somewhat similar to that of Theorem 6.1.

## 9. Conclusion

In this article, we have investigated the recognizability of sets of graphs quite in detail, focusing on the robustness of the notion, which was not immediate since many signatures on graphs can be defined. Although we had in mind sets of graphs, we have proved that embedding graphs in the more general class of relational structures does not alter recognizability. We have proved that the very same structural conditions that equate

VR-equational and HR-equational sets of graphs, also equates HR-recognizability and VR-recognizability.

Summing up, we have defined a number of tools for handling recognizability. Some questions remain to investigate.

- When is it true that a quantifier-free operation preserves recognizability?

Results in this direction have been established in Courcelle [10]. Are they applicable to quantifier-free definable operations? In particular, is it true that the set of disjoint unions of two graphs, one from each of two VR-recognizable sets is VR-recognizable?

- Which quantifier-free definable operations can be added to the signature HR, in such a way that the class of HR-recognizable sets is preserved (as is the case when we extend VR to  $VR^+$ )? The paper by Blumensath and Courcelle [3], which continues the present research, considers unary non qfd operations that can be added to  $VR^+$  and to  $\mathcal{STS}$  while preserving the classes of equational and recognizable sets.
- Our example of an HR-recognizable, not VR-recognizable set of cliques, is based on the weakness of the parallel composition of graphs with sources, i.e., the fact that this operation is not able to split large cliques. Can one find another example, based on a different argument? If one cannot, what does this mean?

We conclude with an observation concerning the finiteness of signatures. Whereas all finite words on a finite alphabet can be generated by this alphabet and only one operation, dealing with finite graphs (by means of grammars, automata and related tools) requires infinite signatures. More precisely, one needs infinitely many operations to generate all finite unlabelled graphs (see Remark 9.1 below). On the other hand, applications to testing graph properties require the consideration of algebras generated by a finite signature. Here is the reason.

Let  $M$  be an  $\mathcal{F}$ -algebra of graphs. If the unique valuation homomorphism  $val_M: T(\mathcal{F}) \rightarrow M$  (which evaluates a term into an element of  $M$ ) is surjective, i.e., if  $\mathcal{F}$  generates  $M$ , then a subset  $L$  of  $M$  is recognizable if and only if  $val_M^{-1}(L)$  is a recognizable set of terms (see Proposition 2.1 and Section 2.3). And the membership of a term in a recognizable set can be verified in linear time by a finite deterministic (tree) automaton. Hence the membership of a graph  $G$  in  $L$  can be checked as follows:

- (1) One must first find some term  $t$  such that  $val_M(t) = G$ ,
- (2) then one checks whether  $t$  belongs to  $val_M^{-1}(L)$ .

The latter step can be done in time proportional to the size of  $t$ , usually no larger than the number of vertices of  $G$ . Although any term  $t$  with value  $G$  gives the correct answer, it may be difficult to find at least one (graph parsing problems may be NP-complete).

Because of this fact many hard problems (in particular if they are expressed in Monadic Second-order logic) can be solved in linear time on sets of graphs of bounded tree-width, and also on sets of graphs of bounded clique-width, provided the graphs are given with appropriate decompositions, see [15,19] or [23]. If the decompositions are not given, one can achieve linear time for graphs of bounded tree-width and  $MS_2$  problems using a result by Bodlaender [4], and polynomial time for graphs of bounded clique-width and  $MS_1$  problems using a result by Oum and Seymour [39].

However, even if  $\mathcal{F}$  is infinite or is finite without generating the set  $M$ , recognizability remains interesting as an algebraic concept, and for every restriction to a finitely generated subset of  $M$ , we are back to the “good” case of a finitely generated algebra.

Finally, we think that infinite signatures can be used for checking graph properties defining recognizable sets. This will not be possible by *finite* tree-automata if the graph algebra is not finitely generated, but it can perhaps be done with automata using “oracles”. An oracle would be a subroutine handling some verifications for big subgraphs that cannot be decomposed by the operations under consideration. This idea needs of course further elaboration.

**Remark 9.1.** We asserted above that finite unlabelled graphs cannot be generated with a finite signature. This is not entirely correct, and we briefly describe here a signature with 6 operations on a 2-sorted algebra which generates, somewhat artificially, all finite graphs (undirected and without loops). These operations have no good behaviour with respect to automata and verification questions, and such an “economical” generation of graphs is useless.

The 2 sorts are  $\mathbf{o}$ , the set of finite graphs equipped with a linear order of their vertex set, and  $\mathbf{u}$ , the set of ordinary, unordered graphs. There is one unary operation of type  $\mathbf{o} \rightarrow \mathbf{u}$ , which forgets the order on the vertex set. All other operations are unary, of type  $\mathbf{o} \rightarrow \mathbf{o}$ : one consists in adding one new vertex, to be the new least element; one adds an (undirected) edge between the two least vertices; one performs a circular shift of the vertices; and one swaps the two least vertices. The three last operations leave the graph unchanged if it has less than 2 vertices. Finally, one adds a 6th, nullary operation, of type  $\mathbf{o}$ : the constant 0, standing for the empty graph with no vertices.

## Appendix A. Equivalences of logical formulas

In this appendix, we discuss some equivalences and transformations of logical formulas which can be used to give upper bounds for the index of congruences considered in this paper, and to complete the proof of the effectiveness of certain notions (e.g. quantifier-free definition schemes).

More specifically, we make precise in what sense we can state, as we do in the body of the paper, that the set of first-order (resp. monadic second-order) formulas over finite sets of relations, constants and free variables, and with a bounded quantification depth, can be considered as finite. Moreover, explicit upper bounds on the size of these finite sets are derived, which can be used to justify the termination of some of our algorithms, and in evaluating their complexity. That these upper bounds have unbounded levels of exponentiation is not unexpected, and even unavoidable by Frick and Grohe [26].

### A.1. Boolean formulas

Let  $p_1, \dots, p_n$  be Boolean variables and let  $B_n$  be the set of Boolean formulas written with these variables. It is well known that  $B_n$  is finite up to logical equivalence. For further reference, we record the following more precise statement.

**Proposition A.1.** *There exists a subset  $B_n^{\text{red}}$  of  $B_n$ , of cardinality  $2^{2^n}$  such that every formula in  $B_n$  can be effectively transformed into an equivalent formula in  $B_n^{\text{red}}$ .*

**Proof.** We let  $B_n^{red}$  be the set of Boolean formulas in disjunctive normal form, where in each disjunct, variables occur at most once and in increasing order, no two disjuncts are equal, and disjuncts are ordered lexicographically. These constraints guarantee the announced cardinality of  $B_n^{red}$ ; the rest of the proof is classical.  $\square$

Of course, the formula in  $B_n^{red}$  equivalent to a given formula, is not always the shortest possible.

### A.2. First-order formulas, semantic equivalence

Let us consider finite sets  $R$  and  $C$ , of relational symbols and of constants (nullary relations, source labels) as in Section 3.1. Recall that, if  $X$  is a finite set,  $FO(R, C, X)$  denotes the set of first-order formulas in the language of  $(R, C)$ -structures, with free variables in  $X$ . For unproved results in this section, we refer the reader to [6].

Several notions of semantic equivalence of formulas can be defined. If  $\varphi, \psi \in FO(R, C, X)$ , say that  $\varphi \equiv \psi$  if for every  $(R, C)$ -structure  $S$  and for every assignment of values in  $S$  to the elements of  $X$ ,  $\varphi$  and  $\psi$  are both true or both false. Say also that  $\varphi \equiv_\omega \psi$  if the same holds for every finite or countable  $(R, C)$ -structure  $S$ , and  $\varphi \equiv_f \psi$  if  $S$  is restricted to being finite.

The equivalences  $\equiv$  and  $\equiv_\omega$  coincide by the Löwenheim-Skolem theorem. Indeed this theorem states that if a closed formula has an infinite model, then it has one of each infinite cardinality: to prove our claim, it suffices to apply it to the formula  $\exists \vec{x} \neg(\varphi(\vec{x}) \Leftrightarrow \psi(\vec{x}))$ . We note that this equivalence cannot be extended to monadic second-order formulas: there exists an MS formula with a unique model, isomorphic to the set of integers  $\mathbb{N}$  with its order.

Each of these three equivalences is known to be undecidable.

The equivalence  $\equiv$  (or  $\equiv_\omega$  since we consider only first-order formulas) is semi-decidable: by Gödel's completeness theorem,  $\varphi \equiv \psi$  if and only if the formula  $\forall \vec{x} (\varphi(\vec{x}) \Leftrightarrow \psi(\vec{x}))$  has a proof, which is a recursively enumerable property.

Trakhtenbrot proved that one cannot decide whether a first-order formula is true in every finite structure, thus proving that  $\equiv_f$  is not decidable. However, the negation of  $\equiv_f$  is semi-decidable: if  $\varphi \not\equiv_f \psi$ , a counter-example can be produced by exploring systematically all finite  $(R, C)$ -structures. This is a proof also that  $\equiv$  and  $\equiv_f$  do not coincide.

### A.3. First-order formulas, a syntactic equivalence

We now describe a syntactic equivalence  $\approx$  on formulas, which refines the semantic equivalences  $\equiv$  and  $\equiv_f$ : that is, if  $\varphi \approx \psi$ , then  $\varphi \equiv \psi$  and  $\varphi \equiv_f \psi$ .

If  $b \in B_n$ , and if  $\varphi_1, \dots, \varphi_n \in FO(R, C, X)$ , we denote by  $b(\varphi_1, \dots, \varphi_n)$  the formula in  $FO(R, C, X)$  obtained by replacing each occurrence of  $p_i$  in  $b$  by  $\varphi_i$ . It is clear that if  $b$  and  $b'$  are equivalent Boolean formulas, then  $b(\varphi_1, \dots, \varphi_n) \equiv b'(\varphi_1, \dots, \varphi_n)$ .

A *Boolean transformation step* consists in replacing in a first-order formula, a sub-formula of the form  $b(\varphi_1, \dots, \varphi_n)$  by the equivalent formula  $b'(\varphi_1, \dots, \varphi_n)$ , where  $b, b' \in B_n$  are equivalent. Then we let  $\varphi \approx \psi$  if  $\varphi$  can be transformed into  $\psi$  by a sequence of Boolean transformation steps and of renamings of bound variables.

It is clear that if  $\varphi \approx \psi$ , then  $\varphi \equiv \psi$ . We want to show that each first-order formula is effectively equivalent to an  $\approx$ -equivalent formula of the same quantifier height, and to give an upper bound on the number of  $\approx$ -equivalence classes of formulas of a given height.

### A.3.1. Quantifier-free formulas

Let  $QF(R, C, X)$  be the set of quantifier-free formulas in  $FO(R, C, X)$ . Such formulas are Boolean combinations of atomic formulas. Let  $Atom(R, C, X)$  be the set of these atomic formulas. Note that each atomic formula is either of the form  $x = y$ , where  $x$  and  $y$  are in  $X \cup C$ , or  $r(x_1, \dots, x_{\rho(r)})$  where  $r$  is a  $\rho(r)$ -ary relation in  $R$  and the  $x_i$  are in  $X \cup C$ . Letting  $n = \text{card}(X)$  and  $c = \text{card}(C)$ , it is easily verified that

$$\text{card}(Atom(R, C, X)) = (n + c)^2 + \sum_{r \in R} (n + c)^{\rho(r)}.$$

We let  $f(R, c, n)$  be this function. Note that if we allow for the (effective) syntactic simplifications of identifying the formulas of the form  $x = x$  with the constant true, and of identifying the formulas  $x = y$  and  $y = x$ , we can lower the value of  $f(R, c, n)$  to  $1 + \frac{1}{2}(n + c)(n + c - 1) + \sum_{r \in R} (n + c)^{\rho(r)}$ .

We then have the following.

**Proposition A.2.** *There exists a subset  $QF^{red}(R, C, X)$  of  $QF(R, C, X)$ , of cardinality  $2^{f(R, c, n)}$ , such that every formula in  $QF(R, C, X)$  can be effectively transformed to an  $\approx$ -equivalent formula in  $QF^{red}(R, C, X)$ .*

**Proof.** By definition of quantifier-free formulas,  $QF(R, C, X)$  is the set of all formulas of the form  $b(\varphi_1, \dots, \varphi_n)$ , where  $b$  is a Boolean formula and the  $\varphi_i$  are atomic formulas. Now let  $QF^{red}(R, C, X)$  be the set of all formulas of the form  $b(\varphi_1, \dots, \varphi_n)$ , where  $b \in B_n^{red}$  and the  $\varphi_i$  are pairwise distinct atomic formulas. The proof of the precise statement is now immediate, using Proposition A.1.  $\square$

**Example A.3.** Let us consider graphs with sources, so that  $R$  consists of a single, binary edge relation. Then  $f(R, c, 0) = 2c^2$  and  $\text{card}(QF^{red}(R, C, \emptyset)) = 2^{2c^2} = q(c)$ . Thus the type equivalence  $\zeta$  (see Section 3.3 and Lemma 3.8) has at most  $2^{q(c)}$  classes in  $\mathcal{GS}(C)$ .

**Remark A.4.** Again, we are not claiming that the set  $QF^{red}(R, C, X)$  is as small as possible. On quantifier-free formulas, the equivalence  $\equiv$  is decidable, because  $\varphi \equiv \psi$  is false if and only if the closed formula  $\exists \vec{x}(\varphi(\vec{x}) \not\equiv \psi(\vec{x}))$  is satisfiable, and the satisfiability problem for existential formulas in prenex normal form is decidable (see [6]). Thus one can modify Proposition A.2 by letting  $QF^{red}(R, C, X)$  be the set of lexicographically minimal formulas in each  $\equiv$ -class: the same statement of Proposition A.2 would then hold with  $\equiv$  instead of  $\approx$ . In particular, the transformation would still be effective, although very inefficient. It is not clear whether the cardinality of the new set of reduced quantifier-free formulas would be significantly smaller.

### A.3.2. Quantifier depth of first-order formulas

Recall that the quantifier depth of a first-order formula is the maximal number of nested quantifiers. If we let  $FO_k(R, C, X)$  be the set of formulas in  $FO(R, C, X)$  of quantifier depth at most  $k$ , a formal definition is as follows:  $FO_0(R, C, X) = QF(R, C, X)$  and, for each  $k \geq 0$ ,  $FO_{k+1}(R, C, X)$  is the set of Boolean combinations of formulas in

$$\begin{aligned} \widehat{FO}_k(R, C, X) = FO_k(R, C, X) \\ \cup \{ \exists y \varphi \mid \varphi \in FO_k(R, C, X \cup \{y\}) \} \\ \cup \{ \forall y \varphi \mid \varphi \in FO_k(R, C, X \cup \{y\}) \}. \end{aligned}$$

Using the same recursion, let us define sets of “reduced” formulas of every quantifier depth. First we fix an enumeration of the countable set of variables. Next, we let  $FO_0^{red}(R, C, X) = QF^{red}(R, C, X)$ . For each  $k \geq 0$ , we then let  $FO_{k+1}^{red}(R, C, X)$  be the set of formulas of the form  $b(\varphi_1, \dots, \varphi_n)$  where  $b \in B_n^{red}$  and the  $\varphi_i$ ’s are in

$$\begin{aligned} \widehat{FO}_k^{red}(R, C, X) = FO_k^{red}(R, C, X) \\ \cup \{ \exists y \varphi \mid \varphi \in FO_k^{red}(R, C, X \cup \{y\}), y \text{ minimal not in } X \} \\ \cup \{ \forall y \varphi \mid \varphi \in FO_k^{red}(R, C, X \cup \{y\}), y \text{ minimal not in } X \}. \end{aligned}$$

**Proposition A.5.** *For each  $k \geq 0$ , the set  $FO_k^{red}(R, C, X)$  is finite. Moreover, every formula in  $FO_k(R, C, X)$  can be effectively transformed to an  $\approx$ -equivalent formula in  $FO_k^{red}(R, C, X)$ .*

**Proof.** Let  $n = \text{card}(X)$  and  $c = \text{card}(C)$ , let  $g(k, R, c, n)$  be the cardinality of  $FO_k^{red}(R, C, X)$ , and let  $h(k, R, c, n)$  be the cardinality of  $\widehat{FO}_k^{red}(R, C, X)$ . It is elementary to verify that these functions can be bounded as follows:

$$\begin{aligned} g(0, R, c, n) &\leq 2^{f(R, c, n)} \text{ and for } k > 0, \\ g(k, R, c, n) &\leq 2^{2^{h(k, R, c, n)}}, \\ h(k, R, c, n) &\leq 3g(k-1, R, c, n+1). \end{aligned}$$

The rest of the proof is immediate, from the recursive definitions.  $\square$

**Remark A.6.** Since there is a procedure to transform each first-order formula into an  $\approx$ -equivalent formula in “reduced form”, we can consider a new equivalence relation on first-order formulas: to yield the same reduced formula. This equivalence is decidable and it refines  $\approx$  (and hence  $\equiv$ ).

**Remark A.7.** In Proposition A.5, we can still consider replacing each formula by the lexicographically least equivalent formula, but this method is not effective, since the equivalence of first-order formulas is not decidable.

#### A.4. Monadic second-order formulas

A very similar analysis can be conducted for monadic second-order formulas of bounded quantifier depth. One difference is that the Löwenheim-Skolem theorem does not hold for these formulas, so the semantic equivalence of formulas based on coincidence on all finite or countable models does not imply coincidence on all models. Moreover, since there is no complete proof systems for such formulas, the equivalences  $\equiv$  and  $\equiv_{\omega}$  are not semi-decidable.

For the rest, one can follow the same techniques as above, to prove the following result. We denote by  $MS_k(R, C, W)$  the set of monadic second-order formulas of quantification depth  $k$  in the language of  $(R, C)$ -structures, with their first- and second-order free variables in  $W$ .

**Proposition A.8.** *For every finite  $R, C, W, k$ , one can construct a finite subset  $MS_k^{red}(R, C, W)$  of  $MS_k(R, C, W)$  such that, for every formula in  $MS_k(R, C, W)$ , one can construct effectively an  $\equiv$ -equivalent formula in  $MS_k^{red}(R, C, W)$ .*

#### References

- [1] K. Abrahamson, M. Fellows, Finite automata, bounded tree-width and well-quasiordering, in: N. Robertson, P. Seymour (Eds.), Graph Structure Theory. Contemp. Math. 147 (1993) 539–564.
- [2] S. Arnborg, B. Courcelle, A. Proskurowski, D. Seese, An algebraic theory of graph reduction, J. ACM 40 (1993) 1134–1164.
- [3] A. Blumensath, B. Courcelle, Recognizability and hypergraph operations using local information, 2004, Submitted.
- [4] H.L. Bodlaender, A linear-time algorithm for finding tree-decompositions of small treewidth, SIAM J. Comput. 25 (1996) 1305–1317.
- [5] B. Bollobas, Extremal Graph Theory, Academic Press, New York, 1978.
- [6] E. Börger, E. Grädel, Yu. Gurevich, The Classical Decision Problems, Springer, Berlin, 1997.
- [7] D.G. Corneil, M. Habib, J.-M. Lanlignel, B. Reed, U. Rotics, Polynomial time recognition of clique-width  $\leq 3$  graphs, Extended abstract. in: G.H. Gonnet, D. Panario, A. Viola (Eds.), LATIN'2000, Lecture Notes in Computer Science, Vol. 1776, 2000, pp. 126–134.
- [8] B. Courcelle, The monadic second-order logic of graphs I: recognizable sets of finite graphs, Inform. Comput. 85 (1990) 12–75.
- [9] B. Courcelle, The monadic second-order logic of graphs VII: graphs as relational structures, Theoret. Comput. Sci. 101 (1992) 3–33.
- [10] B. Courcelle, Recognizable sets of graphs: equivalent definitions and closure properties, Math. Struct. Comput. Sci. 4 (1994) 1–32.
- [11] B. Courcelle, The monadic second-order logic of graphs VI: on several representations of graphs by relational structures, Discrete Appl. Math. 54 (1994) 117–149 Erratum 63 (1995) 199–200.
- [12] B. Courcelle, Structural properties of context-free sets of graphs generated by vertex-replacement, Inform. Comput. 116 (1995) 275–293.
- [13] B. Courcelle, Basic notions of universal algebra for language theory and graph grammars, Theoret. Comput. Sci. 163 (1996) 1–54.
- [14] B. Courcelle, The monadic second-order logic of graphs X: linear orders, Theoret. Comput. Sci. 160 (1996) 87–143.
- [15] B. Courcelle, The expression of graph properties and graph transformations in monadic second order logic, in: G. Rozenberg (Ed.), Handbook of Graph Grammars and Computing by Graph Transformations, Vol. 1, World Scientific, Singapore, 1997, pp. 313–400 (Chapter 5).

- [16] B. Courcelle, The monadic second-order logic of graphs XIV: uniformly sparse graphs and edge set quantifications, *Theoret. Comput. Sci.* 299 (2003) 1–36.
- [17] B. Courcelle, Recognizability of graphs, hypergraphs and relational structures: a survey, in: *DLT 2004*, Lecture Notes in Computer Science, Vol. 3340, Springer, Berlin, 2004, pp. 1–11.
- [18] B. Courcelle, J. Engelfriet, G. Rozenberg, Handle rewriting hypergraph grammars, *J. Comput. System Sci.* 46 (1993) 218–246.
- [19] B. Courcelle, S. Olariu, Upper bounds to the clique-width of graphs, *Discrete Appl. Math.* 101 (2000) 77–114.
- [20] B. Courcelle, J.A. Makowsky, U. Rotics, Linear time solvable optimization problems on graphs of bounded clique-width, *Theory Comput. Systems* 33 (2000) 125–150.
- [21] A. Cournier, M. Habib, A new linear algorithm for modular decomposition, in: S. Tison (Ed.), *CAAP 1994*, Lecture Notes in Computer Science, Vol. 787, Springer, Berlin, 1994, pp. 68–84.
- [22] V. Diekert, *The Book of Traces*, World Scientific, Singapore, 1995.
- [23] R. Downey, M. Fellows, *Parametrized Complexity*, Springer, Berlin, 1997.
- [24] Z. Ésik, Z.L. Németh, Higher dimensional automata, *J. Automata, Languages and Combin.* 9 (2004) 3–29.
- [25] S. Feferman, R. Vaught, The first order properties of products of algebraic systems, *Fundam. Math.* 47 (1959) 57–103.
- [26] M. Frick, M. Grohe, The complexity of first-order and monadic second-order logic revisited, *Ann. Pure Appl. Logic* 130 (2004) 3–31.
- [27] R. Graham, B. Rothschild, J. Spencer, *Ramsey Theory*, Wiley, New York, 1980.
- [28] H.J. Hoogeboom, P. ten Pas, Monadic second-order definable text languages, *Theory Comput. Systems* 30 (1997) 335–354.
- [29] D. Kuske, Towards a language theory for infinite  $N$ -free pomsets, *Theoret. Comput. Sci.* 299 (2003) 347–386.
- [30] D. Kuske, Regular sets of infinite message sequence charts, *Inform. Comput.* 187 (2003) 90–109.
- [31] D. Lapoire, Recognizability equals Monadic Second-Order definability, for sets of graphs of bounded tree-width, in: *STACS 98*, Lecture Notes in Computer Science, Vol. 1373, Springer, Berlin, 1998, pp. 618–628.
- [32] K. Lodaya, P. Weil, Series-parallel languages and the bounded-width property, *Theoret. Comput. Sci.* 237 (2000) 347–380.
- [33] K. Lodaya, P. Weil, Rationality in algebras with a series operation, *Inform. Comput.* 171 (2001) 269–293.
- [34] J. Makowsky, Algorithmic aspects of the Feferman-Vaught Theorem, *Ann. Pure Appl. Logic* 126 (2004) 159–213.
- [35] R. McConnell, J. Spinrad, Linear-time modular decomposition and efficient transitive orientation of comparability graphs, in: *Fifth Annu. ACM-SIAM Symp. on Discrete Algorithms*, ACM, 1994, pp. 536–545.
- [36] R. McConnell, J. Spinrad, Modular decomposition and transitive orientation, *Discrete Math.* 201 (1999) 189–241.
- [37] J. Mezei, J. Wright, Algebraic automata and context-free sets, *Inform. Control* 11 (1967) 3–29.
- [38] R.H. Möhring, F.J. Radermacher, Substitution decomposition for discrete structures and connections with combinatorial optimization, *Ann. Discrete Math.* 19 (1984) 257–356.
- [39] S. Oum, P. Seymour, Approximating clique-width and branch-width, 2004, Submitted.
- [40] D. Perrin, J.-E. Pin, *Infinite Words*, Pure and Applied Mathematics, Vol. 141, Elsevier, Amsterdam, 2004.
- [41] G. Rozenberg, *Handbook of Graph Grammars and Computing by Graph Transformations*, World Scientific, Singapore, 1997.
- [42] S. Shelah, The monadic theory of order, *Ann. Math.* 102 (1975) 379–419.
- [43] W. Thomas, Automata on infinite objects, in: J. van Leeuwen (Ed.), *Handbook of Theoretical Computer Science*, Vol. B, Elsevier, Amsterdam, 1990, pp. 133–192.
- [44] E. Wanke,  $k$ - $NLC$  graphs and polynomial algorithms, *Discrete Appl. Math.* 54 (1994) 251–266.
- [45] W. Wechler, *Universal Algebra for Computer Scientists*, Springer, Berlin, 1992.
- [46] P. Weil, On the logical definability of certain graph and poset languages, *J. Automata, Languages Comput.* 9 (2004) 147–165.
- [47] P. Weil, Algebraic recognizability of languages, in: J. Fiala, V. Koubek, J. Kratochvíl (Eds.), *MFCS 2004*, Lecture Notes in Computer Science, Vol. 3153, Springer, Berlin, 2004, pp. 149–175.