



**HAL**  
open science

## Induced betweenness in order-theoretic trees

Bruno Courcelle

► **To cite this version:**

| Bruno Courcelle. Induced betweenness in order-theoretic trees. 2021. hal-03175157v4

**HAL Id: hal-03175157**

**<https://hal.science/hal-03175157v4>**

Preprint submitted on 29 Nov 2021 (v4), last revised 7 Sep 2022 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Induced betweenness in order-theoretic trees

Bruno Courcelle  
LaBRI, CNRS, Bordeaux University  
courcell@labri.fr

November 30, 2021

**Abstract** : *Betweenness* is an abstract topological notion that has been studied for a long time in different structures. Informally, the ternary relation  $B(x, y, z)$  states that an element  $y$  is *between*  $x$  and  $z$ , in a sense that depends on the considered structure. In a partially ordered set  $(N, \leq)$ ,  $B(x, y, z) : \iff x < y < z \vee z < y < x$ . The corresponding *betweenness structure* is  $(N, B)$ . The class of betweenness structures of linear orders is *first-order definable*; in other words, it is axiomatized by a first-order sentence. That of partial orders is *monadic second-order definable*.

An *order-theoretic tree* is a partial order  $(N, \leq)$  such that, the set of elements larger than any element is linearly ordered, and any two elements have an upper-bound. A rooted tree  $T$  ordered by the ancestor relation is an order-theoretic tree. In an order-theoretic tree, we define  $B(x, y, z)$  to mean that  $x < y < z$  or  $z < y < x$  or  $x < y \leq x \sqcup z$  or  $z < y \leq x \sqcup z$  provided the least upper-bound  $x \sqcup z$  of  $x$  and  $z$  is defined when  $x$  and  $z$  are incomparable. In a previous article, we established that the corresponding class **BO** of betweenness structures is monadic second-order definable. We left as a conjecture that the class **IBO** of *induced substructures* of the structures in **BO** is monadic second-order definable. We prove this conjecture. Our proof uses *partitioned probe cographs* (related to cographs), and their six finite minimal excluded induced subgraphs called their *bounds*. This proof links two apparently unrelated topics: cographs and order-theoretic trees.

However, the class **IBO** has finitely many *bounds*, *i.e.*, minimal excluded finite induced substructures. Hence it is first-order definable. The proof of finiteness uses well-quasi-orders and does not provide the finite list of bounds. Hence, the associated first-order defining sentence is not yet known.

## Introduction

Betweenness is an abstract topological notion that has been studied for a long time in different structures [1, 2, 16, 19, 21, 22]. Informally, the ternary relation of *betweenness*  $B(x, y, z)$  states that an element  $y$  is *between*  $x$  and  $z$  in a sense that depends on the considered structure. In a partially ordered set  $(N, \leq)$ ,  $B(x, y, z) : \iff x < y < z \vee z < y < x$ . The corresponding *betweenness structure* is  $(N, B)$ . The class of betweenness structures of linear orders is *first-order definable*, *i.e.*, is axiomatized by a single first-order sentence; that of partial orders is *monadic second-order definable* [8, 10, 18].

The betweenness structures of certain finite graphs have been studied in [1, 2] and those of trees of various kinds in [4, 6, 7].

An *order-theoretic forest* (an *O-forest* in short) is a partial order  $T = (N, \leq)$  such that  $\{x \in N \mid x \geq y\}$  is linearly ordered for any  $y$ , a notion studied by Fraïssé [13] (under the name of *tree*). It is an *order-theoretic tree* (an *O-tree* in short) if any two elements have an upper-bound. A rooted tree ordered by the ancestor relation is an O-tree. The ordered set of rational numbers is an O-tree in which no node has an immediate ancestor. In previous articles, we used O-trees to define the modular decomposition and the rank-width of countable graphs [4, 9].

The *betweenness structure* of an O-forest  $T = (N, \leq)$  is  $(N, B_T)$  such that

$$B_T(x, y, z) : \iff x \neq y \neq z \neq x \wedge [(x < y \leq x \sqcup z) \vee (z < y \leq x \sqcup z)]$$

where  $x \sqcup z$  denotes the least upper-bound of  $x$  and  $z$ . If  $x \sqcup z$  does not exist, there is no triple  $(x, y, z)$  in  $B_T$ . We denote by **BO** this class of structures. An *induced betweenness* is a induced substructure of such  $(N, B_T)$  where  $T$  can be, equivalently, an O-tree.

In [4, 7], we have axiomatized by *monadic second-order sentences* several classes of betweenness structures related to O-trees, in particular **BO**. We conjectured that the class **IBO** of *induced* betweenness structures in O-trees is monadic second-order definable too. In the present article, we prove a slight weakening of this conjecture by allowing the defining monadic second-order sentence to use the finiteness set predicate  $Fin(X)$ , expressing that a set  $X$  is finite. Furthermore, as we explain below, this class is *first-order definable*, however, the proof does not construct the defining sentence.

The proof of monadic second-order definability uses *partitioned probe cographs*. They are related as follows to *cographs*, the graphs without induced paths  $P_4$ . A *probe cograph* is obtained from a cograph by choosing a subset  $L$  of its vertex set and by removing the edges whose two ends are in  $L$ . In a partitioned probe cograph, we keep track of  $L$  by labelling its vertices. A probe cograph is obtained from a partitioned probe cograph by forgetting such labels. The path  $P_5$  is a probe cograph. The path  $P_6$  is not.

The class of partitioned probe cographs is *hereditary*, *i.e.*, is closed under taking induced subgraphs. A *bound* of such a class  $\mathcal{C}$  is a minimal induced subgraph that is not in  $\mathcal{C}$ , a terminology used by Pouzet [20]. Partitioned probe cographs have six bounds, determined in [17].

The class of probe cographs is also hereditary. As it is *2-well-quasi-ordered* [11, 20], it has finitely many bounds, among which the path  $P_6$ . We exhibit a

few others, without proving that the list is complete. This list is not completely known. We define an algorithm, based on monadic second-order logic and decompositions of the graphs as tree-like unions of complete bipartite graphs that *could* compute them. Unfortunately, this algorithm is intractable.

Our proof links two apparently unrelated notions: betweenness in O-trees and a variant of cographs. To give an intuition about this link, we observe that the cocomparability graph of an O-tree is a graph without induced  $P_4$ , hence, is a possibly infinite cograph (Proposition 3.5).

Finally, we use a result by Pouzet [20] to prove that the class **IBO** of ternary structures has finitely many *bounds*, *i.e.*, minimal excluded induced finite substructures, hence is first-order definable. This proof uses well-quasi-orders of labelled ternary structures, and does not provide the finite list of bounds. Hence, the associated first-order defining sentence is not known. We have no algorithm, even intractable (as for probe cographs), to determine the bounds of **IBO**. We explain this fact by exhibiting a property of clique-width that does not extend from graphs to ternary structures, as we would need.

To summarize, we consider the following hereditary classes of graphs or ternary structures having finitely many bounds:

- cographs, the unique bound is  $P_4$ ;
- partitioned probe cographs, they have six known bounds;
- probe cographs, the finite set is unknown but computable;

**IBO**, the finite set is unknown and we have no algorithm, even intractable, to compute it.

**Summary:** Section 1 reviews partial orders, graphs, clique-width, trees and logic. Section 2 defines and studies probe cographs. Section 3 defines order-theoretic trees. Section 4 proves the two logical characterizations of **IBO**. Section 5 exhibits some bounds of the class of probe cographs. Section 6 states and discusses three open problems.

**Acknowledgement:** I thank Maurice Pouzet for fruitful email exchanges, from which I obtained the FO-definability of the class **IBO**, by a proof based on one of his works. I thank the referees for many useful comments, especially regarding references.

## 1 Definitions and notation

In this article, all graphs, trees, partial orders and, more generally, all relational structures are *countable*, which means finite or countably infinite<sup>1</sup>. We denote by  $[k]$  the set  $\{1, \dots, k\}$ . In some cases, we denote by  $A \uplus B$  the union of two sets  $A$  and  $B$  to stress that they are disjoint.

---

<sup>1</sup>The restriction to countable structures makes it possible to define effective descriptions and to obtain computability results. For example, it is decidable whether two regular join-trees are isomorphic [5], Corollary 3.31.

### Partial orders

For a partial order  $\leq, \subseteq$ , we denote respectively by  $<, \subset$ , the corresponding strict partial order. We write  $x \perp y$  if  $x$  and  $y$  are incomparable for the considered order.

Let  $P = (V, \leq)$  be a partial order. For  $X, Y \subseteq V$ , the notation  $X < Y$  means that  $x < y$  for every  $x \in X$  and  $y \in Y$ . We write  $X < y$  instead of  $X < \{y\}$  and similarly for  $x < Y$ . We use a similar notation for  $\leq$  and  $\perp$ . The least upper-bound of  $x$  and  $y$  is denoted by  $x \sqcup y$  if it exists and is then called their *join*.

If  $X \subseteq V$ , then we define  $N_{\leq}(X) := \{y \in V \mid y \leq X\}$  and similarly for  $N_{<}$ . We have  $N_{\leq}(X) \leq X$  and  $N_{\leq}(\emptyset) = V$ . We define  $L_{\geq}(X) := \{y \in V \mid y \geq X\}$ , and similarly  $L_{>}(X)$ . We write  $L_{\geq}(x)$  (resp.  $L_{\geq}(x, y)$ ) if  $X = \{x\}$  (resp.  $X = \{x, y\}$ ) and similarly for  $L_{>}$ .

An *embedding* of a partial order  $P = (N, \leq)$  into another one  $P' = (N', \leq')$  is an injective mapping  $h : N \rightarrow N'$  such that  $h(x) \leq' h(y)$  if and only if  $x \leq y$ . Hence it is monotone (or isotone). It is a *join-embedding* if  $h(x) \sqcup' h(y) = h(x \sqcup y)$  whenever  $x$  and  $y$  have a join in  $P$ . We write  $P \subseteq_j P'$  if  $N \subseteq N'$  and the inclusion mapping is a join-embedding. If  $P$  and  $P'$  are labelled, then the labels are preserved in embeddings.

### Graphs

Graphs are undirected and simple, which means without loops and parallel (multiple) edges. We denote respectively by  $P_n, C_n, K_n$  a path, a cycle and a clique with  $n$  vertices.

The notation  $u - v$  designates an edge with ends  $u$  and  $v$ . As a property, it means also "there is an edge between  $u$  and  $v$ ". We say then that  $u$  and  $v$  are *adjacent* or *are neighbours*. The notation  $u - v - w - x$  shows the vertices and edges of a path of 4 vertices. The notation  $u - v - w - x - u$  shows the vertices and edges of a 4-cycle.

Induced subgraph inclusion is denoted by  $\subseteq_i$  and  $G[X]$  is the induced subgraph of  $G = (V, E)$  with vertex set  $X \subseteq V$ . Then,  $G - x := G[V - \{x\}]$ .

We denote by  $G \oplus H$  the union of disjoint graphs  $G$  and  $H$ . We define  $G \otimes H$  as  $G \oplus H$  augmented with edges between any vertex of  $G$  and any vertex of  $H$ . We can use the notation  $G_1 \oplus G_2 \oplus \dots \oplus G_n$  because the operation  $\oplus$  is associative, and similarly for  $\otimes$ . We can also use the notation  $\oplus(G_1, G_2, \dots, G_n)$  or  $\otimes(G_1, G_2, \dots, G_n)$  if we consider  $t$  as a rooted tree whose internal vertices are labelled by  $\oplus$  or  $\otimes$ .

The *diameter* of a connected graph is the maximal *distance* between two vertices, *i.e.*, the minimum number of edges of a path between them.

By a *class* of graphs, we mean a set closed under isomorphism.

### Rooted trees

In Graph Theory, a *tree* is a connected graph without cycles. It is convenient to call *nodes* the vertices of a tree because in some proofs, we will discuss simultaneously a graph and an associated tree. A *rooted tree* is a triple  $T = (N_T, E_T, r)$  such that  $(N_T, E_T)$  is a tree and  $r \in N$  is a distinguished node

called its *root*. This tree can be defined from the partial order  $(N_T, \leq_T)$  such that  $x \leq_T y$  if and only if  $y$  is on the path from the *root*  $r$  to  $x$ . In most cases, we will handle a rooted tree  $T$  as a partial order  $(N_T, \leq_T)$ . In Section 3, we will generalize rooted trees into *order-theoretical trees*, defined as partial orders, as done by Fraïssé [13].

A *leaf* is a minimal node and  $L_T$  denotes the set of leaves. The other nodes are *internal*.

If  $x <_T y$ , then  $y$  is an *ancestor* of  $x$ . A node  $x$  is a *son* of  $y$  if  $x <_T y$  and there is no node  $z$  such that  $x <_T z <_T y$ . The *degree* of a node is the number of its sons. A node of degree 0 is thus a leaf.

The *subtree* of a rooted tree  $T = (N, \leq)$  issued from a node  $u$  is  $T/u := (N_{\leq}(u), \leq')$  where  $\leq'$  is the restriction of  $\leq$  to  $N_{\leq}(u)$ .

A *rooted forest*  $F$  is the union of pairwise disjoint rooted trees and  $Rt_F$  denotes the set of roots of its trees.

A finite rooted tree  $T$  can be denoted linearly by  $\tilde{T}$  defined as follows (which is useful in inductive proofs) :

if  $T$  is reduced to  $r$ , then  $\tilde{T} := r$ ,

if  $T$  has root  $r$  and subtrees  $T_1, \dots, T_p$  issued from the sons of the root, then  $\tilde{T} := r(\tilde{T}_1, \dots, \tilde{T}_p)$ .

Any permutation of the sequence  $T_1, \dots, T_p$  defines the same tree because there is no defined order between the sons of a node.

### Relational structures and logic

A *relational structure* is a tuple  $S = (D, R_1, \dots, R_p)$  where  $D$  is a set, its *domain* and  $R_1, \dots, R_p$  are relations of fixed arity. The *signature* of  $S$  is the sequence of arities of the relations  $R_1, \dots, R_p$ . We will consider classes of structures having a fixed signature.

If  $S$  is a relational structure with domain  $N$  and  $X \subseteq N$ , then  $S[X]$  denotes the induced substructure with domain  $X$ , and  $\subseteq_i$  denotes an induced inclusion of relational structures of same signature.

We will use structures  $(N, \leq)$  to describe a partial order, a rooted tree or an order-theoretic forest (defined in Section 3),  $(V, \text{edg})$  to describe an undirected graph with set  $V$  of vertices where  $\text{edg}(x, y)$  means that there is an edge between  $x$  and  $y$ , and  $(N, B)$  for a *betweenness structure* (cf. Introduction and Section 4), where  $B$  is a ternary relation. Additional unary relations will formalize labellings of the elements of  $N$  or  $V$ .

The isomorphism of relational structures is denoted by  $\simeq$ . The isomorphism class of a structure  $S$  is denoted by  $[S]_{\simeq}$ . A set of structures is called a *class* if it is closed under isomorphism. We say that it is *finite* if the set of its isomorphism classes is finite. A class  $\mathcal{C}$  is *hereditary* if it is closed under taking induced substructures.

Properties of structures (and of graphs) will be expressed by *first-order (FO)* or *monadic-second order (MSO)* formulas and sentences. A *sentence* is a formula without free variables. For an example, that a graph has no induced subgraph

isomorphic to a finite graph  $H$  is FO expressible. That a graph is not connected is expressed in its representing structure  $(V, \text{edg})$  by the following MSO-sentence:

$$\exists X[(\exists x.x \in X) \wedge (\exists y.y \notin X) \wedge (\forall x, y.(x \in X \wedge y \notin X \implies \neg \text{edg}(x, y))].$$

The book [10] contains a detailed study of monadic second-order logic.

We will consider classes of countable (which means "finite or countably infinite") structures. Such a class is *MSO* (or *FO*) *definable* if it is the class of countable models of an MSO (or FO)-sentence. It is *uFO definable*, if it is defined by a *universal FO-sentence*, i.e., of the form  $\forall x, y, z, \dots \varphi(x, y, z, \dots)$  where  $\varphi(x, y, z, \dots)$  is quantifier-free.

The finiteness of an arbitrary set  $X$  is not MSO expressible. However it is if some linear order on  $X$  can be defined by MSO-formulas in the case where  $X$  is part of the domain of a structure with a nonempty signature (see Example 1.6 in [7]). This is the case for an example if  $N$  is the set of nodes of a tree of bounded degree. An  $\text{MSO}_{fin}$ -sentence is an MSO-sentence where the *finiteness set predicate*  $\text{Fin}(X)$  expressing that a set  $X$  is finite can be used.

These definitions and the next proposition apply to graphs represented by structures  $(V, \text{edg})$ .

A class of structures  $\mathcal{C}$  is *finitary* if a structure  $S$  belongs to  $\mathcal{C}$  if and only if all its *finite* induced substructures belong to  $\mathcal{C}$ . This implies that  $\mathcal{C}$  is hereditary and characterized by its subclass  $\mathcal{C}_{fin}$  of finite structures. If  $\mathcal{C}$  is finitary and contains an infinite structure, then  $\mathcal{C}_{fin}$  is hereditary but not finitary.

If  $\mathcal{C}$  is finitary, then its *bounds*, forming the class  $\text{Bnd}(\mathcal{C})$ , are the finite structures not in  $\mathcal{C}$  whose proper induced substructures are all in  $\mathcal{C}$ . Then  $\mathcal{C}$  is the class of structures having no induced substructure isomorphic to one in  $\text{Bnd}(\mathcal{C})$ .

A class may be hereditary while having infinitely many bounds. Consider for an example the class of graphs without cycles whose vertices have all degree 2; then each cycle  $C_n$ , where  $n \geq 3$  is a bound of this class.

A routine proof can establish the following.

**Proposition 1.1** : A class of structures is uFO definable if and only if it is finitary and has a finite set of bounds (where finiteness is up to isomorphism). If it is so, its finite structures can be recognized in polynomial time.

## 2 Cographs and related notions

In this section all graphs (and trees) are finite.

**Definition 2.1** : *Cographs*

(a) A graph is a *cograph* if and only if it can be generated from isolated vertices by the operations  $\oplus$  and  $\otimes$ , if and only if it has no induced path  $P_4$ . There are many other characterizations [23]. The family of cographs is hereditary. Its only bound is  $P_4$ .

(b) Cographs can thus be defined by algebraic terms over  $\oplus$  and  $\otimes$  and nullary symbols denoting vertices. For example, the cycle  $a - b - c - d - a$  is

defined by the term  $(a \oplus c) \otimes (b \oplus d)$ . To define it up to isomorphism, that is, without naming the vertices, we will use the term  $(\bullet \oplus \bullet) \otimes (\bullet \oplus \bullet)$ .

(c) We can use the notation  $t_1 \oplus t_2 \oplus \dots \oplus t_n$  because the operation  $\oplus$  is associative, and similarly for  $\otimes$ . We can also use the notation  $\oplus(t_1, t_2, \dots, t_n)$  or  $\oplus(t_1, t_2, \dots, t_n)$ .

(d) The syntactic tree of a term defining a cograph  $G = (V, E)$  is called a  $\{\oplus, \otimes\}$ -tree. It is a rooted tree whose set of leaves is  $V$  and whose internal nodes are of degree at least 2 and labelled by  $\oplus$  or  $\otimes$ .

**Definition 2.2 :** *2-graphs*

A *2-graph* is a graph  $(V, E)$  equipped with a bipartition  $V_1 \uplus V_2$  of its vertex set  $V$ . We will say that  $x \in V_i$  is an *i-vertex*. The *type* of a finite path  $x_1 - x_2 - \dots - x_n$  in a 2-graph is the word  $b_1 b_2 \dots b_n$  over  $\{1, 2\}$  such that  $x_j$  is a  $b_j$ -vertex for each  $j = 1, \dots, n$ .

**Definitions 2.3 :** *Probe cographs*

(a) A *partitioned probe cograph* (a *pp-cograph* in short) is a 2-graph obtained from a cograph  $(V, E)$  by choosing a bipartition  $V_1 \uplus V_2$  of  $V$  and removing the edges between its 1-vertices.

(b) A *probe cograph* (a *p-cograph* in short) is obtained from a pp-cograph by forgetting the bipartition (and the corresponding labelling of its vertices by 1 or 2).

(c) A bipartition of a graph (or its corresponding vertex-labelling by 1 or 2) is *good* if it makes it into a pp-cograph.

(d) Partitioned probe cographs can be defined by terms, similar to those that define cographs, using the operation  $\oplus$  and the operation  $\otimes$  that we redefine as follows for 2-graphs:  $G \otimes H$  is  $G \oplus H$  augmented with all edges between an *i*-vertex of  $G$  and a *j*-vertex of  $H$ , provided *i* and *j* are not both 1. These two operations do not modify the vertex labellings of  $G$  and  $H$ . They are associative. A nullary symbol  $\bullet_i(x)$  defines  $x$  as an isolated *i*-vertex.  $\square$

The path  $P_4 = a - b - c - d$  with labelling of type 1212 is a pp-cograph defined by the term  $\bullet_1(c) \otimes ([\bullet_1(a) \otimes \bullet_2(b)] \oplus \bullet_2(d))$ . To define it up to isomorphism, we can use the term  $\bullet_1 \otimes ([\bullet_1 \otimes \bullet_2] \oplus \bullet_2)$ . Note that  $\bullet_1(x) \otimes \bullet_1(y)$  and  $\bullet_1(x) \oplus \bullet_1(y)$  define the same 2-graph. See also Example 2.5(1).  $\square$

We review some results from [11, 17].

**Proposition 2.4 :** (1) The class of partitioned probe cographs is hereditary. Its bounds are the paths of types 11, 2222, 1222, 2122 or 21212 and the 2-graph  $Q$  defined as the path  $a - b - c - d - e$  of type 12221 augmented with the edge  $b - d$ . Partitioned probe graphs can be recognized in linear time.

(2) The class of probe cographs is hereditary and has finitely many bounds. Its graphs can be recognized in linear time.

An immediate consequence of interest for the present article is that pp-cographs are uFO definable among 2-graphs. The defining sentence is effectively



constructed from the six *known* bounds. Probe cographs are so, but the corresponding uFO sentence is not known, because the complete list of bounds is not either. However, their bounds are definable by a known MSO-sentence obtained from the FO-sentence that defines the pp-cographs. We will discuss these points in Section 5.

**Examples 2.5:** (1) The path  $P_4$  is not a cograph. It has good labellings of types 1212 and 1221. Its labellings of type 1222, 2122 and 2222 are not good.

(2) The labelled path  $P_5 = a - b - c - d - e$  of type 12121 is a pp-cograph defined by the term  $\bullet_1(c) \otimes [(\bullet_1(a) \otimes \bullet_2(b)) \oplus (\bullet_1(e) \otimes \bullet_2(d))]$ . No other labelling of it is good, which follows from Proposition 2.4(1).

(3) The path  $P_6 = a - b - c - d - e - f$  is not a p-cograph. Assume it has a good labelling. The induced path  $a - b - c - d - e$  must have type 12121 and  $f$  must have label 2. But then  $b - c - d - e - f$  has type 21212, which is not possible by (1). It follows that a p-cograph has no induced  $P_6$ . Hence, a connected p-cograph cannot have diameter 5 or more because otherwise, it would contain an induced path  $P_6$ . Furthermore,  $P_6$  is a bound of p-cographs.

(4) A similar proof using (1) shows that the cycle  $C_5$  is a bound of p-cographs. All other graphs having at most 5 vertices are p-cographs. Proposition 5.5 will present some bounds for p-cographs.  $\square$

### 3 Order-theoretic trees

**Definition 3.1:** *Order-theoretic forests and trees.*

(a) An *order-theoretic forest* (an *O-forest* in short) is a partial order  $J = (N, \leq)$  such that, for each  $x \in N$ , called the set of *nodes*, the set  $L_{\geq}(x) := \{y \mid y \geq x\}$  is linearly ordered. An O-forest is an *O-tree* if every two nodes have an upper-bound. An O-tree is a *join-tree*<sup>2</sup> if every two nodes  $x$  and  $y$  have a least upper-bound, denoted by  $x \sqcup y$  and also called their *join*. An O-tree may have no largest node. Its largest node if it exists is called the *root*. If  $x \sqcup y$  and  $y \sqcup z$  are defined, then so is  $x \sqcup z$  and it belongs to  $\{x \sqcup y, y \sqcup z\}$ .

(b) If  $u < w$ , then we say that  $w$  is an *ancestor* of  $u$ .

(c) A *line* in an O-forest  $(N, \leq)$  is a subset  $L$  of  $N$  that is linearly ordered and *convex*, *i.e.*, is such that  $z \in L$  if  $x, y \in L$  and  $x < z < y$ .

(d) A *leaf* is a minimal node. It has *degree* 0; the set of leaves is denoted by  $L_J$ .

(e) A node  $x$  has *degree* 1 if there is  $y < x$  such that every node  $z < x$  is comparable with  $y$ . For finite forests, this is equivalent to the definition in Section 1. If we delete some nodes of degree 1 of an O-forest  $J$ , we obtain a (possibly empty) O-forest  $J'$  that join-embeds into  $J$  (cf. Section 1) because a node of degree 1 is not the join of any two incomparable nodes.  $\square$

<sup>2</sup>We used join-trees to define the modular decomposition and the rank-width of countable graphs [4, 9]. We studied them in algebraic and logical perspectives in [5].

The partial order  $(N_T, \leq_T)$  associated with a rooted tree  $T$  is a join-tree such that  $L_{\geq}(x)$  is finite for each node  $x$ . Conversely, every O-tree having this property is associated in this way with a rooted tree.

**Definition 3.2 :** *Substitutions of lines in O-forests.*

Let  $J = (N, \leq)$  be an O-forest and, for each  $x \in N$ , let  $(A_x, \leq_x)$  be a (possibly empty) linearly ordered set. These sets are assumed to be pairwise disjoint. We let  $J' = J[x \leftarrow A_x; x \in N] := (N', \leq')$  be the partial order such that :

$N'$  is the union of the sets  $A_x$ ,

$u \leq' v$  if and only if either  $u \leq_x v$  or  $u \in A_x \wedge v \in A_y \wedge x < y$ , for some  $x, y$ .

It is an O-forest in which each nonempty set  $A_x$  is a line.

**Definitions 3.3 :** *The join-completion of an O-forest.*

Let  $J = (N, \leq)$  be an O-forest and  $\mathcal{K}$  be the set of upwards closed lines of the form  $L_{\geq}(x, y) := L_{\geq}(x) \cap L_{\geq}(y)$  for all (possibly equal) nodes  $x, y$ . If  $x$  and  $y$  have a join, then  $L_{\geq}(x, y) = L_{\geq}(x \sqcup y)$ . If they have no upper-bound, then  $L_{\geq}(x, y)$  is empty.

The family  $\mathcal{K}$  is countable. We let  $h : N \rightarrow \mathcal{K}$  map  $x$  to  $L_{\geq}(x)$  and  $\widehat{J} := (\mathcal{K}, \supseteq)$ . We call  $\widehat{J}$  the *join-completion of  $J$*  because of the following proposition, stated with these hypotheses and notation.

**Proposition 3.4** [7] : The partially ordered set  $\widehat{J} := (\mathcal{K}, \supseteq)$  is a join-tree and  $h$  is a join-embedding  $J \rightarrow \widehat{J}$ .  $\square$

If we identify  $x \in N$  with  $h(x) := L_{\geq}(x)$ , then  $h$  defines a join-embedding of  $J$  into  $\widehat{J}$ . The join of  $h(x)$  and  $h(y)$  is  $L_{\geq}(x, y)$ .

The following side proposition shows that cographs arise naturally from O-forests. We recall that  $\perp$  denotes incomparability in a partial order.

**Proposition 3.5:** The cocomparability graph  $CC(J) := (N, \perp)$  of a finite forest  $J = (N, \leq)$  is a cograph.

**Proof sketch:** First we prove that the cocomparability graph  $CC(T) = (N, \perp)$  of a finite rooted tree  $T = (N, \leq)$  is a cograph. If  $T = a(T_1, \dots, T_n)$  and  $n \geq 2$ , then  $CC(T) = a \oplus (CC(T_1) \otimes \dots \otimes CC(T_n))$ . If  $n = 1$ , we have  $CC(T) = a \oplus CC(T_1)$ .

If  $J = (N, \leq)$  is a finite forest, it is the disjoint union of rooted trees  $T_1, \dots, T_n$ , then  $CC(J) = CC(T_1) \otimes \dots \otimes CC(T_n)$ .  $\square$

If we define as a cograph any finite or infinite graph without induced path  $P_4$ , then this proposition extends to countable O-forests.

## 4 Betweenness in order-theoretic trees

We will consider ternary structures  $S = (N, B)$ . If  $n > 2$ , the notation  $\neq(x_1, x_2, \dots, x_n)$  means that  $x_1, x_2, \dots, x_n$  are pairwise distinct, hence it abbreviates an FO-formula. If  $n > 3$ , then  $B^+(x_1, x_2, \dots, x_n)$  abbreviates the FO-formula

$$B(x_1, x_2, x_3) \wedge B(x_2, x_3, x_4) \wedge \dots \wedge B(x_{n-2}, x_{n-1}, x_n)$$

and  $A(x_1, x_2, x_3)$  abbreviates

$$B(x_1, x_2, x_3) \vee B(x_2, x_1, x_3) \vee B(x_1, x_3, x_2).$$

**Definitions and background 4.1 :** *Betweenness in O-forests.*

(a) The *betweenness relation* of an O-forest  $J = (N, \leq)$  is the ternary relation  $B_J \subseteq N^3$  such that :

$$B_J(x, y, z) :\iff \neq(x, y, z) \wedge ([x < y \leq x \sqcup z] \vee [z < y \leq x \sqcup z]).$$

We have  $B_J(x, y, z)$  if  $x < y < z$ . If  $x \sqcup z$  is undefined, then  $B_J(x, y, z)$  holds for no triple  $(x, y, z)$ .

We denote by **BO** the class of betweenness structures  $(N, B_J)$  of O-forests  $J = (N, \leq)$ .

(b) The following related classes have been considered in [6, 7].

**IBO** is the class of induced substructures of the structures in **BO**.

**QT** (for *quasi-trees*<sup>3</sup>) is the class of betweenness structures of join-trees.

**IBQT** is the class of induced substructures of structures in **QT**.

We have the following proper inclusions in [7] :

$$\mathbf{BO} \subset \mathbf{IBO}, \mathbf{IBQT} \subset \mathbf{IBO} \text{ and } \mathbf{QT} \subset \mathbf{IBQT} \cap \mathbf{BO}.$$

The classes **IBQT** and **BO** are incomparable, and for finite structures, we have **QT** = **BO**.

(c) The betweenness relation  $B$  of a rooted tree  $T = (N, \leq)$ , (hence  $(N, B) \in \mathbf{QT}$ ) satisfies the following properties for all  $x, y, z, u \in N$ :

$$A1 : B(x, y, z) \Rightarrow \neq(x, y, z).$$

$$A2 : B(x, y, z) \Rightarrow B(z, y, x).$$

$$A3 : B(x, y, z) \Rightarrow \neg B(x, z, y).$$

$$A4 : B(x, y, z) \wedge B(y, z, u) \Rightarrow B^+(x, y, z, u).$$

$$A5 : B(x, y, z) \wedge B(x, u, y) \Rightarrow B^+(x, u, y, z).$$

$$A6 : B(x, y, z) \wedge B(x, u, z) \Rightarrow y = u \vee B^+(x, u, y, z) \vee B^+(x, y, u, z).$$

$$A7 : \neq(x, y, z) \Rightarrow A(x, y, z) \vee \exists w [B(x, w, y) \wedge B(y, w, z) \wedge B(x, w, z)].$$

---

<sup>3</sup>Introduced in [4].

Conversely, every ternary structure satisfying these properties is in **QT** [4]. Hence, the class **QT** is FO-definable. It is not hereditary. Its closure under taking induced substructures, denoted by **IBQT**, is uFO definable by Proposition 2.12 of [7]. It is defined by A1-A6 together with :

$$\text{A8} : \forall u, x, y, z [\neq(u, x, y, z) \wedge B(x, y, z) \wedge \neg A(u, y, z) \Rightarrow B(x, y, u)].$$

The class **BO** is MSO definable [6, 7]. The case of **IBO** was left as a conjecture. We will prove the following two results.

**Theorem 4.2** : (1) The class **IBO** is effectively  $\text{MSO}_{fin}$  definable.  
(2) This class is uFO definable.

Assertion (2) is not effective: we do not know the defining sentence. To the opposite, Assertion (1) is. It entails that the class of bounds of **IBO** is MSO definable (among finite structures). One can prove that  $\text{Bnd}(\mathbf{IBO})$  is finite, but this fact and the knowledge of the defining MSO-sentence are not sufficient to yield an algorithm (see Section 6).

We will consider ternary structures  $(N, B)$  that always satisfy the uFO expressible properties A1-A6. These properties hold in every structure in **IBO** but do not characterize this class (Proposition 3.22 of [7]).

## 4.1 Preliminary results on IBO

**Definition 4.3** : The *Gaifman graph* of a ternary structure  $S = (N, B)$  is the graph  $Gf(S)$  whose vertex set is  $N$  and that has an edge  $u - v$  if and only if  $u$  and  $v$  belong to some triple in  $B$ . We say that  $S$  is *connected* if  $Gf(S)$  is. If it is not, then  $S$  is the disjoint union of the induced structures  $S[X]$  for all connected components  $Gf(S)[X]$  of  $Gf(S)$ .

**Lemma 4.4:** (1) A structure  $S$  is in **IBO** if and only if its connected components are.

(2) If a structure  $S$  in **IBO** is connected, then it is an induced betweenness structure of an O-tree.

**Proof:** (1) The "only if" direction is clear by the definitions. Conversely, assume that each connected component of a ternary structure  $S = (N, B)$  is in **IBO**. For each of them  $S[X]$ , let  $U_X := (M_X, \leq_X)$  be a defining O-forest (we have  $M_X \supseteq X$ ). We let  $\mathbb{N}^R$  be  $\mathbb{N}$  ordered by reversing the natural order. We assume these forests  $U_X$  pairwise disjoint and disjoint from  $\mathbb{N}^R$ . We let  $W$  be the union of  $\mathbb{N}^R$  and the  $U_X$ 's that we order as follows :

$$\begin{aligned} &x \leq_W y \text{ if and only if} \\ &x \leq y \text{ in } \mathbb{N}^R \text{ or } x \leq_X y \text{ for some component } X, \text{ or } x \text{ is in some } M_X \\ &\text{and } y \in \mathbb{N}. \end{aligned}$$

Then  $W$  is an O-tree and  $B = B_W \cap N^3$ .

(2) Let  $S = (N, B)$  be such that  $B = B_U \cap N^3$  for some O-forest  $U = (M, \leq)$ . Let  $M'$  be the union of the lines  $L_{\geq}(x)$  of  $U$  for all  $x \in N$ . Then  $U' := U[M']$  is an O-forest and  $B = B_{U'} \cap N^3$ . We prove that it is an O-tree if furthermore  $S$  is connected. If  $x$  and  $y$  belong to a triple in  $B$ , then they have an upper-bound in  $M'$  by the definition of  $B_U$  and, furthermore, any  $x' \geq x$  and  $y' \geq y$  also have an upper-bound in  $M'$ . Let  $u, v \in M'$ . There is a path  $x_1 - x_2 - \dots - x_n$  in  $Gf(S)$  such that  $u \geq x_1$  and  $v \geq x_n$ . Hence we have  $z_1, z_2, \dots, z_{n-1}$  such that:

- $z_1$  is an upper-bound of  $u$  and  $x_2$ ,
- $z_2$  is an upper-bound of  $z_1$  and  $x_3, \dots$ , and finally
- $z_{n-1}$  is an upper-bound of  $z_{n-2}$  and  $v \geq x_n$ .

We have  $z_{n-1} \geq u$ . Hence,  $U'$  is an O-tree.  $\square$

The converse of Assertion (2) may be false: consider a *star*  $T = (N, \leq)$  with root<sup>4</sup>  $r$  and  $S := (N - \{r\}, B)$  where  $B := B_T[(N - \{r\})]$ . Then,  $S$  is in **IBO**, defined from a tree, but not connected as  $B$  is empty.

**Definition 4.5 :** *Marked join-trees and related notions*

(a) A *marked join-tree* is a 4-tuple  $T = (M, \leq, M_{\oplus}, M_{\otimes})$  such that  $(M, \leq)$  is a join-tree and  $M_{\oplus}, M_{\otimes}$  are disjoint subsets of  $M$  that contain no leaf. We let  $V_T := M - (M_{\oplus} \uplus M_{\otimes})$ . Its *size* is defined as  $|M|$ , the cardinality of  $M$ .

(b) We define the *betweenness relation*  $B_T \subseteq V_T^3$  of  $T$  as follows:

$$B_T(x, y, z) : \iff \neq (x, y, z) \wedge x \sqcup_T z \notin M_{\oplus} \wedge ([x < y \leq x \sqcup_T z] \vee [z < y \leq x \sqcup_T z]).$$

The join  $x \sqcup_T z$  is always defined as  $T$  is a join-tree. We have  $B_T(x, y, z)$  if  $x < y < z$ .

We define the *betweenness structure* of  $T$  as  $S_T := (V_T, B_T)$ . Its Gaifman graph has vertex set  $V_T$ .

(c) If we delete from  $T$  all nodes of degree 1 belonging to  $M_{\oplus} \uplus M_{\otimes}$ , we obtain a marked join-tree having the same betweenness structure and that join-embeds into  $T$  (cf. Definition 3.1(e)). We call *reduced* such a marked join-tree.

(d) If  $M_{\oplus}$  is empty, then  $(V_T, B_T) \in \mathbf{IBQT}$ .

(e) We say that a marked join-tree  $U = (N, \leq, N_{\oplus}, N_{\otimes})$  join-embeds into a marked join-tree  $T = (M, \leq, M_{\oplus}, M_{\otimes})$  if there is a join-embedding of  $(N, \leq)$  into  $(M, \leq)$  that maps  $N_{\oplus}$  to  $M_{\oplus}$  and  $N_{\otimes}$  to  $M_{\otimes}$ .

**Lemma 4.6 :** Let  $T = (M, \leq, M_{\oplus}, M_{\otimes})$  be a marked join-tree.

(1) If  $U = (N, \leq, N_{\oplus}, N_{\otimes})$  join-embeds into  $T$ , then  $B_U = B_T[N \cap V_T]$ .

(2) If  $X \subseteq V_T$  and  $B = B_T[X]$ , then there exists  $U$  as in (1) such that  $B_U = B$ .

---

<sup>4</sup>All other nodes are adjacent to the root.

(3) If  $T_1, \dots, T_n, \dots$  is a sequence of marked join-trees such that  $T_n \subseteq_j T_{n+1}$  and  $T$  is the union of the  $T_n$ 's, then  $B_T$  is the union of the increasing sequence  $B_{T_1} \subseteq_i B_{T_2} \subseteq_i \dots B_{T_n} \subseteq_i \dots$

**Proof :** (1) Since  $U$  join-embeds into  $T$ , if  $x, y \in N \cap V_T$ , then  $x \sqcup_U y = x \sqcup_T y$  and this join belongs to  $M_\oplus$  (resp.  $M_\otimes$ ) if and only if it belongs to  $N_\oplus$  (resp. to  $N_\otimes$ ). The result follows from the definitions.

(2) Let  $T = (M, \leq, M_\oplus, M_\otimes)$  be a marked join-tree and  $N \subseteq M$ . Let us remove from  $T$  all subtrees  $T/u$  that contain no node of  $N$ . We obtain  $U = (N', \leq, N_\oplus, N_\otimes)$ , a marked join-tree that join-embeds into  $T$  and  $V_U = N \cap V_T$ . Hence we have  $B = B_T[N] = B_U$  by (1).

(3) We have  $T_n \subseteq_j T$  for each  $n$ . The result follows.  $\square$

**Proposition 4.7 :** (1) A structure  $S = (N, B)$  is in **IBO** if and only if  $B = B_T$  for a marked join-tree  $T = (M, \leq, M_\oplus, M_\otimes)$  such that  $V_T = N$ .

(2) If  $N$  is finite, then  $T$  can be chosen finite of size at most  $2|N| - 1$ .

**Proof:** (1) "If" direction. Let  $S_T := (V_T, B_T)$  be defined from a marked join-tree  $T = (M, \leq, M_\oplus, M_\otimes)$ . We will construct an O-tree  $U = (W, \leq')$  such that  $M - M_\oplus \subseteq W$  and  $B_U[V_T] = B_T$ .

For each node  $x$  in  $M_\oplus$ , we let  $\mathbb{N}_x^R$  be an isomorphic copy of  $\mathbb{N}$  ordered by reversing the natural ordering. Hence  $\mathbb{N}_x^R$  has no least element. We choose these copies pairwise disjoint and disjoint with  $M$ .

We define  $U := T[x \leftarrow \mathbb{N}_x^R; x \in M_\oplus]$ . It is an O-tree by Definition 3.2 (where substitutions are defined).

If  $x \sqcup_T z \in M_\oplus$ , then  $x$  and  $z$  have no join in  $U$ .

Let  $x, y, z \in V_T$  be such that  $B_T(x, y, z)$  holds. If  $x < y < z$  or  $z < y < x$  in  $T$ , then the same holds in  $U$  and  $B_U(x, y, z)$  holds. Otherwise,  $x$  and  $z$  are incomparable and  $x < y \leq x \sqcup_T z > z$  or  $x < x \sqcup_T z \geq y > z$ . Then,  $x \sqcup_T z$  is either in  $V_T$  or is labelled by  $\otimes$ . In both cases,  $x \sqcup_T z$  is the join of  $x$  and  $z$  in  $U$ . Hence,  $B_U(x, y, z)$  holds.

Conversely, assume that  $x, y, z \in V_T$  and  $B_U(x, y, z)$  holds. If  $x < y < z$  or  $z < y < x$  in  $U$ , then the same holds in  $T$  and  $B_T(x, y, z)$  holds. Otherwise,  $x$  and  $z$  are incomparable and  $x < y \leq x \sqcup_U z > z$  or  $x < x \sqcup_U z \geq y > z$ . Then  $x$  and  $z$  have a join  $m$  in  $T$ . It must be in  $V_T \cup M_\otimes$ , otherwise,  $x \sqcup_U z$  does not exist because it would be the minimal element of  $\mathbb{N}_m^R$ . Hence  $B_T(x, y, z)$  holds. Hence  $S \in \mathbf{IBO}$ .

"Only if" direction. Conversely, assume that  $S = (N, B)$  in **IBO** is defined from an O-tree  $U = (M, \leq)$  such that  $N \subseteq M$  and  $B = B_U \cap N^3$ . We can assume that for every  $y \in M$ , we have  $x \leq y$  for some  $x \in N$ : if this is not the case, we replace  $M$  by the union  $M'$  of the upwards closed lines  $L_{U \geq}(x)$  for all  $x \in N$  and, letting  $U' := (M', \leq)$ , we have  $N \subseteq M'$  and  $B = B_{U'} \cap N^3$ .

Let  $W = (P, \leq)$  be the join-completion of  $U$ , cf Definition 3.3. We label by  $\otimes$  a node in  $M - N$ , and by  $\oplus$  a node in  $P - M$ . These latter nodes have been added to  $U$  in place of missing joins, according to Proposition 3.4.

*Claim :*  $B = B_W$ .

*Proof :*  $B \subseteq B_W$ . Let  $B(x, y, z)$ . If  $x < y < z$  or  $z < y < x$  in  $U$  then the same holds in  $W$  and  $B_W(x, y, z)$  holds.

Otherwise,  $x$  and  $z$  are incomparable and  $x < y \leq x \sqcup z > z$  or  $x < x \sqcup z \geq y > z$  in  $U$ . Then  $x \sqcup z$  is in  $N$  or is labelled by  $\otimes$  in  $W$ . Hence,  $B_W(x, y, z)$  holds. Then  $x \sqcup z$  is not labelled by  $\oplus$ .

Conversely, assume that  $B_W(x, y, z)$  holds. A similar proof establishes that  $B(x, y, z)$  holds.  $\square$

If  $S = (N, B)$  in **IBO** is defined from an O-forest  $U = (M, \leq)$  as opposed to an O-tree, then its connected components are defined by O-trees. For each of them, we have a marked join-tree. We put them together in a marked join-tree with a root labelled by  $\oplus$ . (Similarly to the proof of Lemma 4.4(2)).

(2) Let  $S = (N, B)$  in **IBO** be finite and defined from a marked join-tree  $T = (M, \leq, M_\oplus, M_\otimes)$  such that  $N = V_T$  and  $B = B_T$ . By removing the nodes in  $M_\oplus \cup M_\otimes$  of degree 1, we obtain a reduced marked join-tree that defines  $S$  and has at most  $2|N| - 1$  nodes.  $\square$

**Remark 4.8 :** We observed in Proposition 2.15 of [7] that a finite structure in **IBO** may not be defined from any *finite* O-forest  $U$  (cf. Definition 4.1). Marked join-trees remedy this "defect" and yield Proposition 4.7, a key fact for our proof.

**Example 4.9 :** We consider  $S = (N, B)$  in **IBO** defined from the infinite O-tree on the left of Figure 1 where  $N = \{0, a, a', b, b', c, c', d, d', e, e'\}$  and the dotted line represents  $\mathbb{N}^R$ , without a least node. We have:

- (a)  $B(a', a, 0), B(b', b, 0), B(c', c, 0), B(d', d, 0), B(e', e, 0),$
- (b)  $B^+(a', a, b, b'), B^+(c', c, d, d'), B^+(a', a, e, e'), B^+(b', b, e, e'),$   
 $B^+(c', c, e, e')$  and  $B^+(d', d, e, e')$ .

We do not have  $B^+(a', a, c, c')$  because  $a$  and  $c$  have no join.

The right part shows a finite marked join-tree  $T = (M, \leq, M_\oplus, M_\otimes)$  where  $z$  has been added as join of  $x$  and  $y$  and the nodes  $2, 3, \dots, n, \dots$  of degree 1 above  $z$  have been deleted (cf. Definition 3.1 for the degree).

We have  $M = N \cup \{1, x, y, z\}, M_\oplus = \{z\}, M_\otimes = \{1, x, y\}$  and  $B_T = B$ .  $\square$

**Proposition 4.10 :** The class **IBO** is finitary, that is,  $S$  is in **IBO** if and only if each of its finite induced substructures is.

**Proof :** The "only if" direction is clear as, by its definition, the class **IBO** is hereditary, *i.e.*, closed under taking induced substructures.

"If" direction. First, some observations. If  $S = (V_T, B)$  is defined from a marked join-tree  $T = (N, \leq, N_\oplus, N_\otimes)$  and  $S' \subseteq_i S$ , then the restriction  $T'$  of  $T$  to  $\{x \in N \mid x \geq y \text{ for some } y \in V_T\}$  is a marked join-tree that defines  $S'$ . By reducing it (Definition 4.5(c)), we get  $T'' \subseteq_j T$  that defines  $S'$ .

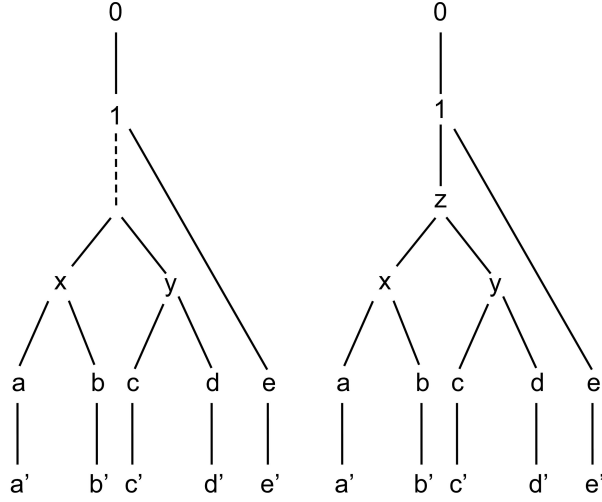


Figure 1: See Example 4.9. In the O-tree to the left, we have  $0 > 1 > 2 > 3 > \dots > n > \dots$  above  $x$  and  $y$ .

By Proposition 4.7, each finite structure  $S = (N, B)$  in **IBO** of size  $m = |N|$  is defined by marked join-trees of size at most  $2m - 1$ . We let  $J(S)$  be the finite set of all such join-trees, up to isomorphism.

For proving the statement, we let  $S = (N, B)$  be infinite. It is the union of an increasing sequence  $S_1 \subset_i S_2 \subset_i \dots \subset_i S_n \subset_i \dots$  of finite induced substructures that we assume to be in **IBO**.

We will use the following version of Koenig's Lemma. Let  $A_1, A_2, \dots, A_n, \dots$  be an infinite sequence of pairwise disjoint finite sets, and  $A$  be their union. Let  $R \subseteq A \times A$  be such that for every  $b$  in  $A_n, n > 1$ , there is  $a \in A_{n-1}$  such that  $(a, b) \in R$ . Then, there exists an infinite sequence  $a_1, \dots, a_n, \dots$  such that  $(a_{n-1}, a_n) \in R$  for each  $n > 1$ .

The finite sets  $J(S_n)$  are pairwise disjoint. We define

$$R := \{(T, T') \mid T \in J(S_{n-1}), T' \in J(S_n), n > 1 \text{ and } T \subset_j T'\}.$$

It follows from Lemma 4.6(2) that if  $T' \in J(S_n)$  and  $n > 1$ , we have  $(T, T') \in R$  for some  $T \in J(S_{n-1})$ .

Hence, there is an infinite sequence of marked join-trees

$$T_1 \subset_j T_2 \subset_j \dots \subset_j T_n \subset_j \dots \text{ such that } T_n \in J(S_n) \text{ for each } n.$$

By Lemma 4.6(3), their union is a marked join-tree  $T$  such that  $T_n \subset_j T$  for each  $n$ . We obtain an increasing sequence of finite marked join-trees whose union is a marked join-tree that defines  $S$ . Hence  $S \in \mathbf{IBO}$ .  $\square$

The proof of Theorem 4.2(1) reduces to that of the following proposition.

**Proposition 4.11:** There is an MSO-sentence that characterizes the finite connected structures in **IBO** among the finite ternary structures.



**Proof of Theorem 4.2(1)**, assuming proved Proposition 4.11 :

Let  $\varphi$  be an MSO-sentence such that, for every finite ternary structure  $S = (N, B)$  :

$S \models \varphi$  if and only if  $S$  is connected and belongs to **IBO**.

Consider the MSO<sub>fin</sub> sentence  $\psi$  :

$$\forall X. (\gamma(X) \wedge \text{Fin}(X) \implies \varphi[X]),$$

where  $\gamma(X)$  expresses that  $X$  is connected in the Gaifman graph  $Gf(S)$  and  $\varphi[X]$  is the relativization of  $\varphi$  to  $X$ .

*Relativizing* a sentence to a set, here  $X$ , is a classical construction in logic, see *e.g.* [10], Section 5.2.1 for monadic second-order logic. If  $S = (N, B)$  is a ternary structure and  $X \subseteq N$ , then  $S \models \varphi[X]$  if and only if  $S[X] \models \varphi$ .

We prove that  $S \models \psi$  if and only if  $S$  is in **IBO**.

If  $S$  is in **IBO**, then every induced substructure  $S[X]$ , in particular every finite and connected one satisfies  $\varphi$ , hence  $S \models \psi$ .

Conversely, assume that  $S \models \psi$ . Let  $X$  be a finite subset of  $N$ . If it is connected in  $Gf(S)$ , then  $\varphi[X]$  holds hence  $S[X]$  is in **IBO**. Otherwise, it is a disjoint union of connected sets in  $Gf(S)$ . For each of them, say  $Y$ , the validity of  $\psi$  implies that  $\varphi[Y]$  holds,  $S[Y]$  is in **IBO** and so are  $S[X]$  by Lemma 4.4(1) and  $S$  by Proposition 4.10.  $\square$

## 4.2 Proof of Proposition 4.11

Proposition 4.11 is the main technical result. We will only handle finite objects: graphs, rooted trees, rooted forests and structures  $(N, B)$ . All trees and forests will be rooted, defined as partial orders  $(N, \leq)$  and simply called trees and forests. We need some more definitions.

**Definition 4.12** : *Forests compatible with a ternary relation.*

A rooted forest  $T = (N, \leq_T)$  is *compatible with a relation*  $B \subseteq N^3$  satisfying Axioms A1-A6 (Definition 4.1) if, for all  $x, y, z \in N$  :

- (i) if  $B(x, y, z)$  holds, then  $x <_T y$  or  $z <_T y$ ,
- (ii) if  $B(x, y, z)$  and  $x <_T y >_T z$  hold, then  $y = x \sqcup_T z$ .
- (iii) if  $x <_T z$ , then  $B(x, y, z)$  holds if and only if  $x <_T y <_T z$ .  $\square$

**Lemma 4.13** : Let  $S = (N, B) \in \mathbf{IBO}$  be finite, connected and defined from a finite reduced marked tree<sup>5</sup>  $U = (N \uplus N_{\oplus} \uplus N_{\otimes}, \leq_U, N_{\oplus}, N_{\otimes})$ .

(1) Then  $T := U[N] = (N, \leq_T)$  is a finite forest compatible with  $B$ , where  $\leq_T$  is the restriction of  $\leq_U$  to  $N$ .

---

<sup>5</sup>Every finite tree is a join-tree.

(2) The order  $\leq_T$  is FO definable in the structure  $(N, B, R)$  where  $R$  is the set of roots of  $T$ , *i.e.* of maximal elements with respect to  $\leq_T$ .  $\square$

The forest  $T$  is not necessarily a tree because the root of  $U$  need not be in  $N$ . This root cannot be labelled by  $\oplus$ , otherwise  $S$  is not connected (we exclude the trivial case where  $N$  is singleton).

**Proof :** (1) Let  $S, T, U$  as in the statement.

(i) If  $B(x, y, z)$  holds, then:

either  $x <_U y <_U z$  or  $z <_U y <_U x$ ,  
or  $x \sqcup_U z \wedge [(x <_U y \leq_U x \sqcup_U z) \vee (z <_U y \leq_U x \sqcup_U z)]$ ,  
where in the latter case,  $x \sqcup_U z \in N \cup N_\otimes$ .

In all cases, we have  $x <_U y$  or  $z <_U y$ , hence  $x <_T y$  or  $z <_T y$ .

(ii) If  $B(x, y, z)$  and  $x <_T y >_T z$  hold, then the above description of  $B(x, y, z)$  shows that  $y \leq_U x \sqcup_U z$ . As we have  $x <_U y >_U z$ , we must have  $y = x \sqcup_U z$ . If  $y$  is not  $x \sqcup_U z$ , we have  $m \in N$  such that  $x < m$  and  $z < m < y$  in  $T$  and in  $U$ . But then  $y$  is not the join of  $x$  and  $z$  in  $U$  and  $y = x \sqcup_U z$ .

(iii) Clear from the definitions because  $\leq_T$  is the restriction of  $\leq_U$  to  $N$ .

(2) If  $R = \{r\}$ , then  $x \leq_T y$  if and only if  $x = y$  or  $y = r$  or  $B(x, y, r)$  holds. Otherwise, the root of  $U$  is in  $N_\otimes$  and has degree at least 2. Let  $x$  and  $y$  be not in  $R$ .

*Claim :* (a) If  $r \in R$ , we have  $x <_T r$  if and only if  $B(x, r, r')$  holds for some  $r' \in R$ .

(b) We have  $x <_T y$  if and only if  $B(x, y, r)$  holds for some  $r \in R$ .

*Proof:* (a) Assume that  $x <_T r$ . There is  $r' \in R$  such that  $r \sqcup_U r'$  has label  $\otimes$ . Hence  $B(x, r, r')$  holds.

Conversely, if  $B(x, r, r')$  holds for some  $r' \in R$ , we have  $x <_T r$  or  $r' <_T r$  because  $T$  is compatible with  $B$ . As  $r$  and  $r'$  are different and are distinct roots of  $T$ , they are incomparable and we have  $x <_T r$ .

(b) If  $x <_T y$ , we have  $x <_T y <_T r$  for some  $r \in R$ . Hence  $B(x, y, r)$  holds since  $T$  is compatible with  $B$ .

Conversely, if  $B(x, y, r)$  holds for some  $r \in R$ , then, we have  $x <_T y$  or  $r <_T y$ . The latter is not possible as  $r$  is a root.  $\square$

Let  $\psi(R, x, y)$  be the following FO formula (an FO formula may have free set variables and use atomic formulas  $x \in X$ ):

$$\begin{aligned} x = y \vee [x \neq y \wedge \exists r. (R = \{r\} \wedge [y = r \vee B(x, y, r)])] \\ \vee [x \neq y \wedge \exists r, r' \in R. (y = r \wedge B(x, y, r'))] \\ \vee [x \neq y \wedge y \notin R \wedge \exists r \in R. B(x, y, r)]. \end{aligned}$$

By the claim, it defines  $x \leq_T y$  since  $R$  is the set of roots of  $T$ .

We let  $\varphi(R)$  be the FO formula relative to ternary structures  $S = (N, B)$  expressing the following:

" $R \subseteq N$ , the binary relation  $x \leq y$  on  $N$  defined by  $S \models \psi(R, x, y)$  is a partial order and  $T := (N, \leq)$  is a forest that is compatible with  $B$  and whose set of roots is  $R$ ".  $\square$

**Proposition 4.14 :** Let  $S = (N, B)$  be finite and satisfy properties A1-A6.

(1) For every  $R \subseteq N$  such that  $S \models \varphi(R)$ , if we let  $\leq$  be defined by  $\psi(R, x, y)$ , then  $T := (N, \leq)$  is a forest compatible with  $B$ .

(2) Every forest  $T := U[N]$  defined from a finite marked tree  $U = (N \uplus N_{\oplus} \uplus N_{\otimes}, \leq_U, N_{\oplus}, N_{\otimes})$  such that  $B_U = B$  is described by the formulas  $\varphi(R)$  and  $\psi(R, x, y)$ .

**Proof :** The first assertion follows from the definition of  $\varphi$ . The second one follows from Lemma 4.13(2).  $\square$

All forests  $T$  compatible with  $B$  of potential interest for checking that  $S$  is in **IBO** can be described in terms of their sets of roots  $R$  by the existential MSO-formulas  $\exists R. \varphi(R)$  and  $\psi(R, x, y)$ .

We will construct MSO-formulas to "check" that a "guessed" forest  $T$  satisfies additional requirements implying that  $T := U[N]$  for some finite marked tree  $U$  witnessing that  $S \in \mathbf{IBO}$ .

In some  $T$  that has been "guessed", we will insert (if possible) finitely many nodes labelled by  $\oplus$  or  $\otimes$ , so as to make it into the desired marked tree  $U$ . We will insert nodes in  $T$  in the following cases:

(1) If  $a' < a$ ,  $b \perp a$ ,  $B(a', a, b)$  holds and there is no  $x$  in  $T$  such that  $\{a, b\} < x$  and  $B(a, x, b)$  holds, then we insert  $a \sqcup_U b$  labelled by  $\otimes$  such that  $a \sqcup_U b < m$  where  $m$  is any upper-bound in  $T$  of  $\{a, b\}$ .

(2) If  $a' < a$ ,  $b \perp a$  but  $B(a', a, b)$  does not hold, then we insert  $a \sqcup_U b$  as above labelled by  $\oplus$ .

In Case (1)  $a$  and  $b$  may have a join  $m$  in  $T$  but we need to insert a "new join"  $a \sqcup_U b < m$ .

**Example 4.15 :** The left part of Figure 2 shows a tree  $T$  constructed from the following facts relative to  $N := \{a', a, b, c', c, d, e', e, 1, 0\}$  :

- (a)  $B^+(a', a, 1, 0), B(b, 1, 0), B^+(c', c, 1, 0), B(d, 1, 0), B^+(e', e, 1, 0),$
- (b)  $B^+(a', a, c, c'), B(a', a, d), B(b, c, c'),$
- (c)  $B^+(a', a, 1, e, e'), B^+(b, 1, e, e'), B^+(c', c, 1, e, e')$  and  $B^+(d, 1, e, e')$ .

Facts (b) indicate the need of joins  $a \sqcup c$ ,  $a \sqcup d$ , and  $b \sqcup c$  labelled by  $\otimes$  in the marked tree  $U$  to be constructed. These joins are all equal to  $x$  in the tree in the middle of Figure 2. The absence of facts  $B(a', a, b)$  and  $B(c', c, d)$  indicates the need of  $\oplus$ -labelled joins  $y$  and  $z$ , respectively between  $a, b$  and  $x$ , and between  $c, d$  and  $x$ . However, no triple in  $B$  necessitates that  $b$  and  $d$  have a  $\otimes$ -labelled join. A corresponding marked tree is shown in the middle of Figure 2.

Consider now  $N' := \{a', a, b, d, e', e, 1, 0\} \subseteq N$  and  $B' := B[N']$ . We get a marked tree for  $B'$  by deleting  $c, c'$  and  $z$  from the previous one. However,

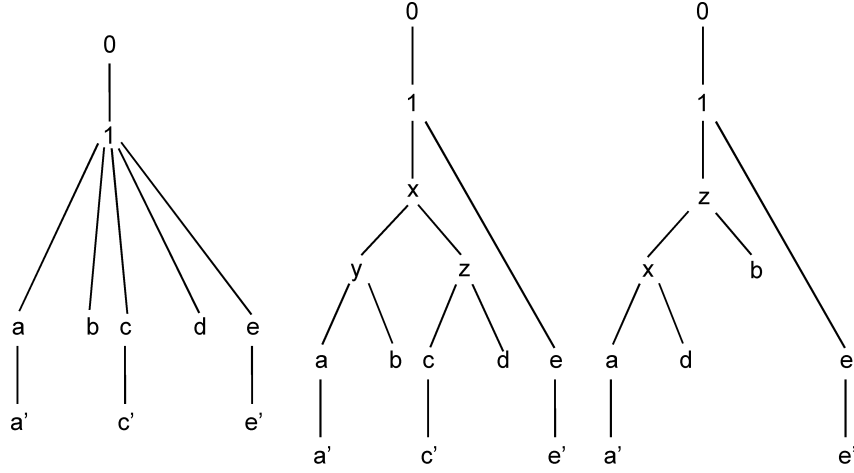


Figure 2: Example 4.15

another one is shown to the right, that defines  $(N', B')$  where  $x$  is labelled by  $\otimes$  and  $z$  by  $\oplus$ .  $\square$

We need more definitions. We let  $B, T, U$  and  $R$ , be as in Lemma 4.13.

**Definitions 4.16 :** *Cographs and pp-cographs defined from  $T$  and, either  $U$  or  $B$ .*

(a) For each node  $x$  of  $T$  with sons  $y_1, \dots, y_s, s \geq 2$ , we define  $y_i \sim_x y_j$  if and only if  $i = j$  or  $y_i \sqcup_U y_j \neq x$  so that this join has label  $\otimes$  or  $\oplus$ . It is clear that  $\sim_x$  is an equivalence relation.

We have  $y_i \sim_x y_j$  if and only if  $B(y_i, x, y_j)$  does not hold, by (ii) of compatibility, Definition 4.12.

(b) For each class  $C$  of the equivalence relation  $\sim_x$ , we define  $G_{x,C}$  as the 2-graph  $(C, E, C_1, C_2)$  such that:

- $y \in C_2$  if and only if  $y' < y$  for some  $y' \in N$ ,
- $y - z$  is an edge<sup>6</sup> if and only if  $y$  or  $z$  is in  $C_2$  and  $y \sqcup_U z$  has label  $\otimes$ .

There are no edges between vertices in  $C_1$ . Hence  $G_{x,C}$  is a pp-cograph. We obtain a cograph if we add edges  $y - z$  such that  $y$  and  $z$  are in  $C_1$  and  $y \sqcup_U z$  has label  $\otimes$ .

(c) Let  $R = \{r_1, \dots, r_p\}, p \geq 2$ . We let  $G_{root}$  be the 2-graph  $(R, E, R_1, R_2)$  defined as  $G_{x,C}$  above, where  $R$  replaces  $C$  and  $y \in R_2$  if and only if  $y' < y$  for some  $y' \in N$ . It is also a pp-cograph.  $\square$

<sup>6</sup> $E$  denotes the set of edges.

**Lemma 4.17 :** The edges  $y - z$  of  $G_{x,C}$  and  $G_{root}$  are characterized by the FO formula  $(\exists y' < y. B(y', y, z)) \vee (\exists z' < z. B(z', z, y))$ .

**Proof :** Consider  $x \in N$  having sons  $y$  and  $z$  in a class  $C$  of  $\sim_x$ .

Let  $y - z$  be an edge of  $G_{x,C}$  such that  $y \in C_2$  and  $y \sqcup_U z$  has label  $\otimes$ . Then  $B_U(y', y, z)$  holds for all  $y' < y$  and so does  $B(y', y, z)$  as  $B = B_U$ .

Conversely, if  $y' < y \wedge B(y', y, z)$  holds, then the join  $y \sqcup_U z$  must have label  $\otimes$  or be in  $N$ . But in the latter case, it must be  $x$  as  $y$  and  $z$  are sons of  $x$ . Hence, we have  $B(y, x, z)$  but then, we do not have  $y \sim_x z$ . Hence,  $y - z$  is an edge of  $G_{x,C}$ .

The proof is similar for  $G_{root}$ . The join  $y \sqcup_U z$  cannot be in  $N$  as  $y, z$  are distinct roots.  $\square$

It follows that the 2-graphs  $G_{root}$  and  $G_{x,C}$  can be defined from  $B$  and  $T$  only, without using  $U$  that we are actually looking for. Furthermore, they can be described by FO formulas in the structure  $(N, B, R)$ .

The formulas  $\varphi(R)$  and  $\psi(R, x, y)$  are defined before Proposition 4.14.

**Proposition 4.18 :** Let  $S = (N, B)$  and  $T = (N, \leq)$  be defined by the formula  $\psi(R, x, y)$  from some  $R$  satisfying  $\varphi(R)$ .

(1) There exists a marked tree  $U \supseteq T$  such that  $B = B_U$  if and only if the 2-graphs  $G_{root}$  and  $G_{x,C}$  are pp-cographs.

(2) This condition is FO expressible in the structure  $(N, B, \leq)$ .

**Proof:** (1) The "only if" direction follows from the previous constructions.

Conversely, assume that each 2-graph  $G_{x,C}$  (determined solely from  $T$  and  $B$  by Lemma 4.17) as in the statement is a pp-cograph. By adding some edges between its 1-vertices, we can get a cograph  $H_{x,C} \supseteq G_{x,C}$ . It is defined by an  $\{\oplus, \otimes\}$ -tree  $t_{x,C}$  (Definition 2.1(c)), a tree whose internal nodes are labelled by  $\oplus$  or  $\otimes$  and whose set of leaves is  $C$ .

Similarly, if  $T$  has several roots and  $G_{root}$  is a pp-cograph, there is a cograph  $H_{root} \supseteq G_{root}$  defined by a  $\{\oplus, \otimes\}$ -tree  $t_{root}$  whose set of leaves is  $R$ .

By inserting in  $T$  the internal nodes of  $t_{x,C}$  between  $x$  and the nodes in  $C$ , for all relevant pairs  $(C, x)$ , and those of  $t_{root}$  above the roots of  $T$ , we get a marked tree  $U$  such that  $U \supseteq T = U[N]$  and  $B = B_U$ .

This can be formalized as follows. By bottom-up induction, we define marked trees  $T_x$  and  $T_{x,C}$  for each  $x$  in  $N$  and equivalence class  $C$  of the relation  $\sim_x$ . We assume that the trees  $t_{x,C}$  and  $t_{root}$  are pairwise disjoint.

(a) If  $x$  is a leaf, then  $T_x := x$ . There is no set  $C$  to consider.

Otherwise,  $T_x := x(\dots, T_{x,C}, \dots)$  where the list covers all equivalence classes  $C$  of  $\sim_x$ . (We use the linear notation of finite rooted trees defined in Section 1).

(b) If  $C = \{y\}$ , then  $T_{x,C} := T_y$ .

Otherwise, we use the  $\{\oplus, \otimes\}$ -tree  $t_{x,C}$  to define  $T_{x,C} := t_{x,C}[\dots, y \leftarrow T_y/y, \dots]$ , denoting the simultaneous substitution in  $t_{x,C}$  of  $T_y$  for each  $y \in C$  (it is a leaf of  $t_{x,C}$ ).

(c) To complete the construction, we define  $U := T_r$  if  $T$  is a tree with root  $r \in N$ . Otherwise,  $U := t_{root}[\dots, r \leftarrow T_r, \dots]$  denoting the substitution in  $t_{root}$  of  $T_r$  for each leaf  $r \in R$  (it is a leaf of  $t_{root}$ ).

It is clear that  $U$  is a marked tree  $(N \uplus N_{\oplus} \uplus N_{\otimes}, \leq_U, N_{\oplus}, N_{\otimes})$  and that  $T = U[N]$ .

*Claim* :  $B_U = B$ .

*Proof* : Note that  $\leq_T$  is the restriction of  $\leq_U$  to  $N$ .

If  $x$  and  $z$  are comparable, then  $(x, y, z) \in B_U$  if and only if  $x <_U y <_U z$  if and only if  $x <_T y <_T z$  if and only if  $(x, y, z) \in B$  since  $T$  is compatible with  $B$ .

We now assume  $x \perp z$  and  $(x, y, z) \in B$ .

Let  $u := x \sqcup_U z$ . By the compatibility of  $T$  with  $B$  (point (ii)), we have  $x <_T y$  or  $z <_T y$ .

(a) If  $u \in N$ , then  $u = x \sqcup_T z$ . Again by compatibility (point (i)), we do not have  $u <_T y$ . Hence, we have  $x <_T y \leq_T u >_T z$  or  $z <_T y \leq_T u >_T x$ . The same inequalities hold with  $\leq_U$  hence  $(x, y, z) \in B_U$ .

(b) Otherwise,  $u$  has label  $\oplus$  or  $\otimes$ . Let  $x'$  be maximal in  $N$  such that  $x \leq x'$  and  $z'$  be similar for  $z$ .

(b.1) If  $x'$  and  $z'$  have no upper-bound in  $T$ , they are distinct roots and  $u$  is an internal node of  $G_{root}$ .

As noted above, we have  $x <_T y$  or  $z <_T y$ . Assume the first, *w.l.o.g.*. Then  $x <_T y \leq_T x'$ .

If  $u$  has label  $\otimes$  then  $B_U(x, y, z)$  holds by the definition of  $B_U$ .

(If  $u$  has label  $\oplus$  then  $B_U(x, y, z)$  does not hold, but the definition of the edges of  $G_{root}$  gives that  $B_U(x, y, z)$  does not hold.)

(b.2) If  $x'$  and  $z'$  have a least upper-bound  $m$  in  $T$ , then,  $u <_U m$ . We have two cases:

*Case 1* :  $B(x, m, z)$  holds. We cannot have  $y >_T m$ , hence, we have  $x <_T y \leq_T m >_T z$  or  $z <_T y \leq_T m >_T x$ . The same inequalities hold with  $\leq_U$ , hence  $(x, y, z) \in B_U$ .

*Case 2* : if  $B(x, m, z)$  does not hold. Then  $x' \sim_m z'$  (we cannot have  $B(x', m, z')$ ) and so  $x'$  and  $z'$  belong to a same class  $C$  of  $\sim_m$ . Then we use the same argument as above with  $G_{m,C}$  instead of  $G_{root}$ .

The proof that  $B_U \subseteq B$  is similar.  $\square$

This completes the proof of Assertion (1).

(2) The following facts can be expressed in the structure  $(N, \leq, B)$  such that  $S = (N, B)$  satisfies A1-A6 and  $T = (N, \leq)$  is a forest compatible with  $B$  by MSO-formulas that are easy to write explicitly:

$\alpha(R, R_1, R_2)$  :  $R$  is the set of root of  $T$ , it is not singleton and  $(R_1, R_2)$  is its partition defined in Definition 4.16(c).

$\beta(x, y, z)$  :  $y < x \wedge z < x \wedge y \sim_x z$ , ( $y$  and  $z$  are sons of  $x$  in  $T$ ) cf. Definition 4.16(a).

$\gamma(x, C, C_1, C_2)$  :  $C$  is a set of sons of  $x$  and an equivalence class of  $\sim_x$ ,  $(C_1, C_2)$  is its partition defined in Definition 4.16(b).

$\eta(R, y, z)$  :  $y - z$  is an edge of  $G_{root}$ .

$\eta'(x, C, y, z) : y - z$  is an edge of  $G_{x,C}$ .

$\pi(R) : R$  is not singleton and  $G_{root}$  is a pp-cograph (we use  $\alpha$  and  $\eta$ ).

$\pi'(x, C, y, z) : G_{x,C}$  is well-defined and is a pp-cograph (we use  $\gamma$  and  $\eta'$ ).

It is MSO expressible in  $(N, B, \leq)$  by Proposition 2.4(1) whether the 2-graphs  $G_{root}$  and  $G_{x,C}$  are all pp-cographs. The condition of Assertion (1) is thus MSO expressible in the structure  $(N, B, \leq)$  by an MSO-sentence  $\mu$ .  $\square$

**Proof of Proposition 4.11:** We prove that an MSO-sentence can characterize the finite connected structures in **IBO** among the finite ternary structures. There is an MSO-sentence  $\chi$  expressing that a ternary structure  $S = (N, B)$  is connected and satisfies A1-A6. The sentence over  $S = (N, B)$  defined as  $\exists R.(\varphi(R) \wedge \mu'(R))$  where  $\mu'$  translates  $\mu$  (of Proposition 4.18(2)) by using  $\psi(R, x, y)$  to define  $\leq$  expresses well that  $S$  is in **IBO** by Proposition 4.18(1). Hence,  $\chi \wedge \exists R.(\varphi(R) \wedge \mu'(R))$  is the desired sentence.  $\square$

### 4.3 Well-quasi-orderings and finite sets of bounds

We recall definitions and a result from [20].

**Definitions 4.19 :**

(a) If  $\mathcal{C}$  is a hereditary class of finite structures  $S = (N, R_1, \dots, R_p)$ , if  $m$  is the maximal arity of a relation  $R_i$ , we denote by  $\mathcal{U}(\mathcal{C})$  the class of structures  $(N, R_1, \dots, R_p, U_1, \dots, U_{2m-1})$  for all  $S$  in  $\mathcal{C}$ , where  $U_1, \dots, U_{2m-1}$  are unary relations, hence that denote subsets of  $N$ .

(b) The class  $\mathcal{U}(\mathcal{C})$  is *well-quasi-ordered* (implicitly by induced inclusion) if, for every infinite sequence  $S_1, S_2, \dots$  of structures in this class, there are  $n < m$  such that  $S_n$  is isomorphic to an induced substructure of  $S_m$ , which we denote by  $S_n \subseteq_{i\sim} S_m$ .  $\square$

With these hypotheses and notation, Corollary 2.4 of [20] (also Theorem<sup>7</sup> 13.2.3 of [13]) states the following. (The class of bounds  $Bnd(\mathcal{C})$  is defined in Section 1. Its finiteness is up to isomorphism.)

**Theorem 4.20 :** If  $\mathcal{C}$  is hereditary and  $\mathcal{U}(\mathcal{C})$  is well-quasi-ordered, then  $Bnd(\mathcal{C})$  is finite.

The structures in  $\mathcal{U}(\mathbf{IBO})$  are of the form  $(N, B, U_1, \dots, U_5)$  for  $(N, B)$  in **IBO**. Let  $\mathbf{T}$  be the set of finite structures  $T = (N \uplus N_{\oplus} \uplus N_{\otimes}, \leq, N_{\oplus}, N_{\otimes}, U_1, \dots, U_5)$  such that  $(N \uplus N_{\oplus} \uplus N_{\otimes}, \leq, N_{\oplus}, N_{\otimes})$  is a marked tree (Definition 4.5(a)),  $U_1, \dots, U_5$  are subsets of  $N$  and  $|N_{\oplus} \uplus N_{\otimes}| < |N|$ .

Proposition 4.7(2) shows that every  $S$  in  $\mathcal{U}(\mathbf{IBO})$  is defined from a marked tree  $T$  belonging to  $\mathbf{T}$ . We denote then  $S = S(T)$ . Precisely,  $S(T) = (N, B_{T'}, U_1,$

<sup>7</sup>Fraïssé states the result with  $2m$  instead of  $2m - 1$  but translates from French the proof by Pouzet.

...,  $U_5$ ) where  $T'$  is the marked tree  $T = (N \uplus N_{\oplus} \uplus N_{\otimes}, \leq, N_{\oplus}, N_{\otimes})$ , cf. Definition 4.5(b).

*Fact* : If  $T, T'$  are in  $\mathbf{T}$  and  $T \subseteq_{j\sim} T'$ , then  $S(T) \subseteq_{i\sim} S(T')$ .

It is a corollary of Lemma 4.6(1).

**Proposition 4.21** : The class of finite structures in  $\mathcal{U}(\mathbf{IBO})$  is well-quasi-ordered.

**Proof** : Let  $S_1, S_2, \dots$  be an infinite sequence of finite structures in  $\mathcal{U}(\mathbf{IBO})$ . For each  $S_n$ , we let  $T_n$  in  $\mathbf{T}$  be such that  $S(T_n) = S_n$ . By Kruskal's Theorem,  $T_n \subseteq_{j\sim} T_m$  for some  $n < m$ . The above fact yields  $S_n \subseteq_{i\sim} S_m$ .  $\square$

**Theorem 4.2(2)** : The class  $\mathbf{IBO}$  has finitely many bounds.

**Proof** : The hereditary class of finite structures belonging to  $\mathbf{IBO}$  has finitely many bounds by Proposition 4.21 and Theorem 4.20. The result holds by Proposition 4.10.  $\square$

**Remark 4.22** : We recall from Definition 4.1 that  $\mathbf{IBQT}$  is the class of induced betweenness of join-trees. It is a proper subclass of  $\mathbf{IBO}$ . The structures in  $\mathbf{IBQT}$  are defined from marked join-trees  $T = (N \uplus N_{\oplus} \uplus N_{\otimes}, \leq, N_{\oplus}, N_{\otimes})$  such that  $N_{\oplus}$  is empty (Definition 4.5(d)). The proof of Theorem 4.2(2) shows that  $Bnd(\mathbf{IBQT})$  is finite, hence that  $\mathbf{IBQT}$  is FO definable, without constructing the defining sentence. The FO definability of  $\mathbf{IBQT}$  is Theorem 3.1 of [7], where the defining FO-sentence is the conjunction of Conditions A1-A6 and A8 of Definition 4.1.

In Section 6, we will explain why computing the bounds of  $\mathbf{IBO}$  is even harder than computing those of probe cographs.

## 5 Clique-width and the bounds of probe cographs

We first review clique-width, then, we discuss some properties of probe cographs, in view of determining their bounds.

**Definition 5.1** : *Clique-width*.

(a) Graphs are built with the help of vertex labels (in addition to the labels of 2-graphs). Each vertex has a *label* in a set  $L$ . The nullary symbol  $\mathbf{a}(x)$  where  $a \in L$ , denotes the isolated vertex  $x$  labelled by  $a$ . The operations are the union  $\oplus$  of disjoint graphs (it does not modify labels), the unary operation  $add_{a,b}$  for  $a, b \in L, b \neq a$ , that adds to a graph an edge between each  $a$ -labelled vertex and each  $b$ -labelled vertex (unless they are already adjacent), the unary operation  $relab_{a \rightarrow b}$  that changes every vertex label  $a$  into  $b$ .

(b) A term over the above defined operations is *well-formed* if no two occurrences of nullary symbols denote the same vertex (so that the graphs defined by



two arguments of any operation  $\oplus$  are disjoint). We call them the *clique-width terms*. Each term  $t$  denotes a vertex labelled graph  $\mathbf{val}(t)$  whose vertices are those specified by the nullary symbols of  $t$ . Its *width* is the number of labels that occur in  $t$ . The *clique-width* of a graph  $G$  without labels from  $L$  (but possibly with labels from another set like  $\{1, 2\}$ ), denoted by  $cwd(G)$ , is the least width of a term  $t$  that denotes some vertex labelling of  $G$ .

(c) Clique-width terms may contain redundancies: for example, we have  $add_{a,b}(add_{c,d}(add_{a,b}(G))) = add_{c,d}(add_{a,b}(G))$  and  $relab_{a \rightarrow b}(relab_{a \rightarrow c}(G)) = relab_{a \rightarrow c}(G)$  for every graph  $G$ . It follows that each graph of clique-width at most  $k$  is defined by infinitely many terms written with a fixed set  $L$  of  $k$  labels. However, one can "normalize" these terms so as to avoid redundancies. This is done in Proposition 2.121 of [10]. Let us call *normal* such a term. Then, each graph of clique-width at most  $k$  is defined by finitely many normal terms using the labels in  $L := [k]$ . Furthermore, the set  $N_k$  of normal terms with labels in  $[k]$  is recognizable by a finite automaton, see [10].  $\square$

**Proposition 5.2 :** The maximal clique-width of a probe cograph is 4.

**Proof :** The upper-bound, observed in [11], is easy to establish. The bound 4 is reached by the probe cograph defined by the term (cf. Definition 2.3(c)):

$$[\bullet_1(1) \oplus (\bullet_1(2) \otimes \bullet_2(7)) \oplus (\bullet_1(3) \otimes \bullet_2(8))] \otimes \\ [\bullet_1(4) \oplus (\bullet_1(5) \otimes \bullet_2(9)) \oplus (\bullet_1(6) \otimes \bullet_2(10))]$$

where the vertices 1,...,6 are 1-vertices and the vertices 7 to 10 are 2-vertices. It has clique-width<sup>8</sup> 4.  $\square$

**Proposition 5.3:** Apart from  $P_6$ , the finitely many bounds of probe cographs have diameter at most 4 and clique-width at most 8. They are connected and MSO definable.

**Proof :** Since the class of probe cographs is closed under disjoint union, its bounds are connected.

We have observed that  $P_6$  of diameter 5 is a bound. Any other graph of diameter at least 5 has an induced path  $P_6$ , hence is not a bound.

If a graph has  $G - x$  has clique-width  $k$ , then  $G$  has clique-width at most  $2k$  [14]. Hence, as probe cographs have clique-width at most 4, their bounds have clique-width at most 8.

If  $\mathcal{C}$  is a hereditary class of finite graphs, then its bounds form the class:

$$Bnd(\mathcal{C}) := \{G \mid G \notin \mathcal{C} \text{ and } G - x \in \mathcal{C} \text{ for each vertex } x \text{ of } G\}.$$

If  $\mathcal{C}$  is defined by an MSO-sentence  $\theta$ , then  $Bnd(\mathcal{C})$  is defined by the MSO-sentence :

$$-\theta \wedge \forall X. ("X \text{ is the set of all vertices minus one"} \implies \theta[X])$$

<sup>8</sup>The verification has been done by using the software TRAG [12] that is accessible on-line. It is based on [15].

By Proposition 2.4(1) the class of pp-cographs is FO-definable. Hence, the class of probe cographs is MSO-definable: an existential set quantification is useful to guess a good labelling of the given graph. The corresponding MSO-sentence is known from the knowledge of the bounds of pp-cographs. However, the class of probe cographs is FO-definable by Proposition 2.4(2), but we do not know the corresponding sentence as the bounds of probe cographs are not completely known.  $\square$

**Theorem 5.4 :** There is an algorithm that can compute the finitely many bounds of the class of probe cographs. An upper-bound to their sizes is computable.

**Proof sketch :** By Proposition 5.3, we can construct effectively an MSO-sentence  $\xi$  that defines the class  $\mathcal{B}$  of bounds of probe cographs among finite graphs. By Theorem 6.35 of [10] or an algebraic version of it in terms of recognizable sets (Corollary 5.59), one can build a finite automaton  $\mathcal{A}$  that recognizes the set of normal terms of width at most 8 that define the graphs in  $\mathcal{B}$ . Then  $L(\mathcal{A})$  is finite as we know that  $\mathcal{B}_{\simeq}$  is. However, several terms in  $L(\mathcal{A})$  may define isomorphic graphs. As  $L(\mathcal{A})$  is finite, one can list its elements and thus the graphs it defines after removing isomorphic duplicates.

The MSO-sentence  $\xi$  can be replaced by  $\xi \wedge \delta$  where  $\delta$  is the MSO-sentence expressing that a graph is connected and has a diameter at most 4. We obtain in this way a more restrictive set  $L(\mathcal{A})$  without missing any graph in  $\mathcal{B}$  except  $P_6$ , but we know it.

*Pumping lemmas* are classical tools of language theory by which one can bound the sizes of the terms of a finite recognizable set defined by a given finite automaton without listing them. However the obtained bound would be ridiculous huge.  $\square$

This decision procedure is actually intractable, because of the complexity of the sentence  $\xi$  and the size of the corresponding automaton, that needs to handle clique-width terms with 8 labels.

#### *Some bounds of probe cographs*

By a *bound*, we mean a bound of probe cographs, hence, a graph in  $\mathcal{B}$ .

We denote by  $\overline{G}$  the edge-complement of a graph  $G$ . By *substituting an edge* (i.e., the graph  $K_2$ ) to a vertex  $a$  of a graph  $G$ , we obtain the graph denoted by  $G[a \leftarrow K_2]$ . Its vertex  $a$  is replaced by the edge  $a_1 - a_2$  and any edge  $a - x$  of  $G$  is replaced by the two edges  $a_1 - x$  and  $a_2 - x$ . Note that  $G[a \leftarrow K_2] - a_1$  is isomorphic to  $G$ . We have  $G[a \leftarrow K_2][b \leftarrow K_2] = G[b \leftarrow K_2][a \leftarrow K_2]$  if  $b \neq a$ . It is the result of the *simultaneous substitution* of  $K_2$  for  $a$  and  $b$ , denoted by  $G[a \leftarrow K_2, b \leftarrow K_2]$ .

**Proposition 5.5 :** The following graphs, all of clique-width 3, are in  $\mathcal{B}$  :

- (1) The standard graphs  $C_5, P_6$  and  $C_6$ .
- (2) The graphs  $\overline{C_6}, D$  and  $\overline{D}$  derived from  $C_6$ , see Figure 3.

(3) Four graphs obtained by substituting an edge to one or two vertices of a path  $P_4$  or  $P_5$ . See Figure 4.

(4) Two graphs obtained from the *house*  $H$  by substituting edges as in (3). See Figure 5.

(5) Two graphs obtained as in (4) from two p-cographs with 5 vertices. See Figure 6.

**Proof:** The proofs are based on the following observations (cf. Example 2.5) :

the only good labellings of  $P_4$  are 1212, 2121 and 1221,

the only good labelling of  $P_5$  is 12121,

the two good labellings of the "house"  $H$  with vertices  $\{a, b, c, d, e\}$  and top vertex  $c$ , cf. Figure 5, label by 1 either  $c$  and  $d$ , or  $c$  and  $b$ ;

the other vertices are labelled by 2,

every good labelling of a graph is good for its induced subgraphs.

(1) See Examples 2.5 for  $C_5$  and  $P_6$ . Let  $C_6 = a-b-c-d-e-f-a$ . Assume for a contradiction that it has a good labelling. By removing  $a$ , we get  $P_5$ , hence  $b$  and  $f$  must be labelled by 1, hence  $a$  must be labelled by 2. Similarly,  $b$  must be labelled by 2. As  $P_5$  is a p-cograph,  $C_6$  is a bound. Hence,  $C_5, P_6$  and  $C_6$  are bounds.

(2) The graph  $\overline{C_6}$  is shown in Figure 3. By removing  $f$ , we get the "house"  $H$  with top vertex  $c$  that should be labelled by 1, so that either  $b$  or  $d$  should be labelled by 1. Hence,  $f$  must be labelled by 2, but it is the top vertex of the house  $\overline{C_6} - c$ . Hence,  $\overline{C_6}$  is not a p-cograph. However, it is a bound.

The graph  $D$  is obtained from  $C_6 = a-b-c-d-e-f-a$  by adding the edge  $c-f$ . By removing  $f$ , we see that  $a, c$  and  $e$  should be labelled by 1, and  $b$  by 2. If we remove  $c$ , we get a path  $P_5$  of type 21212 which is not good.

From the graph  $\overline{D}$  shown in Figure 3, we get two "houses" by removing either  $c$  or  $f$ . Both vertices should be labelled by 1, so that none of the others can be labelled by 1.

(3) The path  $P_5 = a-b-c-d-e$  has a unique good labelling of type 12121. If we substitute  $K_2$  for any of  $a, c$  or  $e$ , we obtain a bound as the two vertices of the substituted edge cannot be both labelled by 1. We obtain only two non-isomorphic bounds, shown in the top part of Figure 4, because substituting an edge to  $a$  and  $e$  give isomorphic graphs.

The path  $P_4 = a-b-c-d$  has three good labellings of types 1212, 2121 and 1221. For each good labelling, at least one vertex in  $\{a, b\}$ , in  $\{a, d\}$  and in  $\{c, d\}$  must be labelled by 1. It follows that  $P_4[a \leftarrow K_2, b \leftarrow K_2]$  and  $P_4[a \leftarrow K_2, d \leftarrow K_2]$  shown in the bottom part of Figure 4 are bounds as one checks easily. So is  $P_4[c \leftarrow K_2, d \leftarrow K_2]$  isomorphic to the first one. We obtain two non-isomorphic bounds.

(4) Every good labelling of the "house"  $H$  shown to the left of Figure 5, must label  $c$  by 1 and, either  $b$  or  $d$ , by 1. We obtain the two bounds  $H[c \leftarrow K_2]$  and  $H[b \leftarrow K_2, d \leftarrow K_2]$  shown in Figure 4.

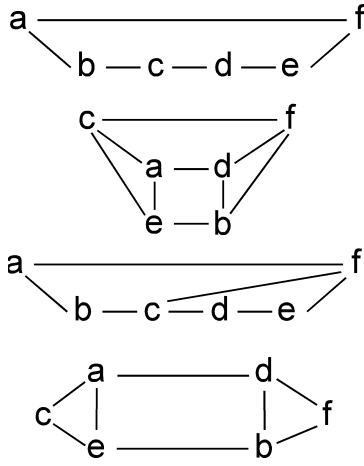


Figure 3: The bounds  $C_6$ ,  $\overline{C_6}$ ,  $D$  and  $\overline{D}$  of Proposition 5.5(1,2).

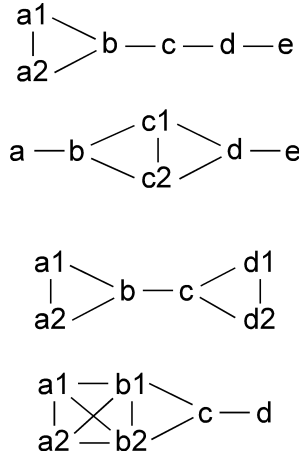


Figure 4: The bounds of Proposition 5.5(3).

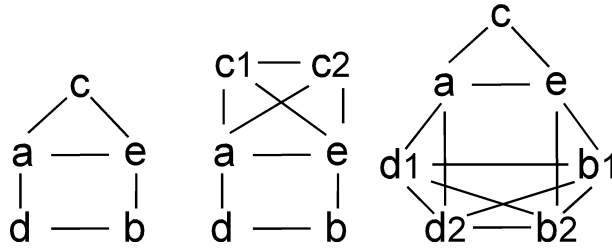


Figure 5: The "house" to the left (it is a p-cograph), and the two bounds of Proposition 5.5(4).

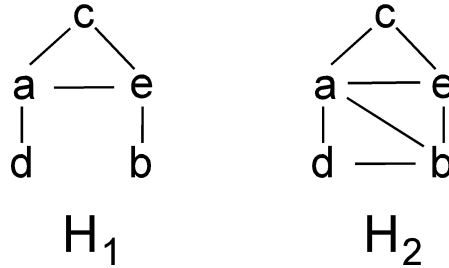


Figure 6: Two p-cographs used in Proposition 5.5(5)

(5) Among the connected graphs with 5 vertices that are not cographs we have are  $C_5$  and  $H$ . Figure 6 shows two others,  $H_1$  and  $H_2$ .

Every good labelling of  $H_1$ , labels  $b$  or  $d$  (or both) by 1. Hence  $H_1[b \leftarrow K_2, d \leftarrow K_2]$  is not a p-cograph. It is a bound because  $H_1[b \leftarrow K_2]$  and  $H_1[d \leftarrow K_2]$  are p-cographs.

Similarly, we get from  $H_2$ , the bound  $H_2[c \leftarrow K_2, d \leftarrow K_2]$ .

The graphs  $C_5, P_6$  and  $C_6$  are known to have clique-width 3. That  $H, \overline{C_6}, D$  and  $\overline{D}$  have clique-width 3 can be checked with [12] or proved directly. All other bounds of (3),(4) and (5) are obtained by substituting  $K_2$  to vertices of graphs of clique-width 3. Hence, they have clique-width 3.  $\square$

We do not know any graph of clique-width 4 or more that is a bound of probe cographs. Hence, for now, we are far from the upper-bound 8 of Proposition 5.3.

## 6 Open problems

**Problem 6.1** : Determine the set of bounds of probe cographs. What can be said about them in addition to what is stated in Proposition 5.3 ?

Monadic second-order logic does not help for effective computations because of the huge sizes of the automata constructed from MSO-sentences.

**Problem 6.2** : Does there exist a *monadic second-order transduction*, cf. Chapter 7 of [10], that transforms a finite ternary structure assumed to be in **IBO** into a finite marked tree defining it?

In [7], we studied four classes of betweenness structures (cf. Definition 4.1) : **QT**, **IBQT**, **BO** and **IBO**. Each betweenness structure  $S = (N, B)$  is defined from a labelled O-tree, say  $T = (M, \leq, N_{\oplus}, N_{\otimes})$ . This description covers all cases, although labels are useless in some cases.

The question is whether some witnessing O-tree  $T$  can be defined by MSO-formulas in the given structure  $S$ , in technical words, by an MSO-transduction. An FO-transduction exists for **QT** and MSO-transductions exist for **IBQT** and **BO** [7].

Theorem 4.2(1) establishes that the class **IBO** is  $\text{MSO}_{fin}$ -definable, without building an associated MSO-transduction. We recall that an MSO-transduction transforms a structure with  $n$  elements into one with at most  $kn$  elements for some fixed  $k$ . As a finite structure in **IBO** having  $n$  elements can be defined from a marked tree with at most  $2n - 1$  nodes, it is not hopeless to look for such a transduction. An intermediate result seems necessary: to find an MSO-transduction that constructs, from a pp-cograph, a term that defines it. Such a transduction exists for countable cographs given with an auxiliary linear order [3, 9]. Such an auxiliary order would be useful, even perhaps necessary.

**Problem 6.3** : Determine the set of bounds of **IBO**.

We have presently no result similar to Theorem 5.4 because ternary structures do not share certain good properties of graphs, as we now explain.

First we observe that, since we have an effective MSO characterization of the finite structures in **IBO** by Theorem 4.2(1), we have one of the set of their bounds because the proof for graphs (Proposition 5.3) extends to relational structures.

The proof of Theorem 5.4 uses the fact that the bounds of probe cographs have clique-width  $\leq 8$  (even if this upper-bound is overestimated). We miss a corresponding fact for  $\text{Bnd}(\mathbf{IBO})$ . First because there is no really convenient notion of clique-width for ternary structures. However, we can replace the property "the graphs of  $\mathcal{C}$  have clique-width at most  $k$ " by "the structures of  $\mathcal{C}$  are all in  $\tau(\mathbf{Trees})$  for some MSO-transduction  $\tau$ " where **Trees** is the class of finite rooted trees, and say that  $\mathcal{C}$  is *tree-definable*. Bounded clique-width is equivalent for a class of graphs to tree-definability ([10], Chapter 7). Furthermore, the computability results for classes of graphs of bounded clique-width hold for tree-definable classes of structures.

For a set  $\mathcal{C}$  of structures of a fixed signature, we let  $\mathcal{C}^+$  be the set of structures  $S$  with domain  $N$  such that  $S[N - x]$  is in  $\mathcal{C}$  for some  $x$  in  $N$ . If  $\mathcal{C}$  is a set of graphs of clique-width at most  $k$ , then the graphs in  $\mathcal{C}^+$  have clique-width at most  $2k$  (see [14]), which is used to obtain the upper-bound 8 to the clique-width of the bounds of probe cographs in Proposition 5.3. However, this fact

does not extend to tree-definable classes of structures as we prove below. Hence, we cannot extend Theorem 5.4 to the computation of  $Bnd(\mathbf{IBO})$ .

**Proposition 6.4** : There is a tree-definable set of ternary structures  $\mathcal{C} \subseteq \mathbf{IBO}$  such that  $\mathcal{C}^+$  is not tree-definable.

**Proof** : We will use results from [10]. Let  $S_G = ([n], B_G)$  where  $B_G$  consists of the triples  $(1, i, j)$  and  $(j, i, 1)$  for the edges  $i - j$  of some graph  $G = (\{2, 3, \dots, n\}, E)$  and  $i < j$ . Then  $S_G$  is a ternary structure that satisfies Properties A1-A6. It is in  $\mathcal{C}^+$  where  $\mathcal{C}$  is the set of trivial ternary structures  $(\{2, 3, \dots, n\}, \emptyset)$ , obviously in  $\mathbf{IBO}$ . There is an MSO-transduction  $\theta$  that transforms each structure  $S_G$  into  $G$  : it deletes 1 and replaces the triples  $(1, i, j)$  and  $(j, i, 1)$  in  $B_G$  by  $(i, j)$  and  $(j, i)$ , thus defining  $E$ .

It is clear that  $\mathcal{C}$  is tree-definable. If  $\mathcal{C}^+$  would be, that is, if  $\mathcal{C}^+ \subseteq \tau(\mathbf{Trees})$  for some MSO transduction  $\tau$ , then the MSO transduction  $\theta \circ \tau$  would produce all finite graphs from  $\mathbf{Trees}$  (up to isomorphism), hence, all graphs would have clique-width bounded by a fixed value ([10], Chapter 7), which is false.  $\square$

## References

- [1] M. Changat, P. Narasimha-Shenoi, and G. Seethakuttyamma, Betweenness in graphs: A short survey on shortest and induced path betweenness, *AKCE International Journal of Graphs and Combinatorics*, **16** (2019) 96-109. (Available on ScienceDirect.com, <https://doi.org/10.1016/j.akcej.2018.06.007>)
- [2] V. Chvatal, Antimatroids, betweenness, convexity, in *Research Trends in Combinatorial Optimization*, Springer, 2008, pp. 57-64.
- [3] B. Courcelle, The monadic second-order logic of graphs X: Linear orderings. *Theor. Comput. Sci.* **160** (1996) 87-143.
- [4] B. Courcelle, Several notions of rank-width for countable graphs, *J. Comb. Theory, Ser. B.* **123** (2017) 186-214.
- [5] B. Courcelle, Algebraic and logical descriptions of generalized trees, *Logical Methods in Computer Science* **13** (2017) Issue 3.
- [6] B. Courcelle, Betweenness in order-theoretic trees, in *Fields of Logic and Computation III, Lecture Notes in Computer Science* **12180** (2020) 79-94.
- [7] B. Courcelle, Axiomatizations of betweenness in order-theoretic trees, *Logical Methods in Computer Science* **17** (2021) Issue 1, pp. 11:1-11-42.
- [8] B. Courcelle, Betweenness of partial orders, *Theoretical Informatics and Applications*, **54** (2020), see <https://www.rairo-ita.org/articles/ita/abs/2020/01/ita200028/ita200028.html>.

- [9] B. Courcelle and C. Delhommé, The modular decomposition of countable graphs. Definition and construction in monadic second-order logic. *Theor. Comput. Sci.* **394** (2008) 1-38.
- [10] B. Courcelle and J. Engelfriet, *Graph structure and monadic second-order logic, a language theoretic approach*, Cambridge University Press, 2012.
- [11] J. Daligault, M. Rao and S. Thomassé, Well-quasi-order of relabel functions. *Order* **27** (2010) 301-315.
- [12] I. Durand and M. Raskin, On line software TRAG, <http://trag.labri.fr>
- [13] R. Fraïssé, *Theory of relations, Studies in Logic*, Volume **145**, North-Holland, 2000.
- [14] F. Gurski, The behaviour of clique-width under graph operations and graph transformations. *Theory Comput. Syst.* **60** (2017) 346-376.
- [15] M. Heule and S. Szeider, A SAT approach to clique-width. *ACM Trans. Comput. Log.* **16** (2015) 24:1-24:27.
- [16] E. V. Huntington and J. R. Kline, Sets of independent postulates for betweenness, *Transactions of the A.M.S.*, **18** (1917) 301-325.
- [17] V. B. Le and H. de Ridder, Characterisations and linear-time recognition of probe cographs. *Graph-Theoretic Concepts in Computer Science*, (WG 2007), *Lecture Notes in Computer Science* **4769** (2007) 226-237.
- [18] J. Lihova, Strict-order betweenness, *Acta Univ. M. Belii Ser. Math.* **8** (2000) 27-33. Available from <https://actamath.savbb.sk/acta0804.shtml>.
- [19] E. Pitcher and M. F. Smiley, Transitivity of betweenness, *Transactions of the A.M.S.*, **52** (1942) 95-114.
- [20] M. Pouzet, Un bel ordre d'abritement et ses rapports avec les bornes d'une multirelation, *Comptes Rendus Académie des Sciences, Série A*, **274** (1972) 1677-1680.
- [21] M. Sholander, Trees, lattices, order, and betweenness, *Proceedings of the A.M.S.* , **3** (1952) 369-381.
- [22] M. Sholander, Medians and betweenness, *Proceedings of the A.M.S.*, **5** (1954) 801-807.
- [23] Wikipedia, Cographs, <https://en.wikipedia.org/wiki/Cograph>