A study of algorithms relating distributive lattices, median graphs, and Formal Concept Analysis

Alain Gély, Miguel Couceiro, Laurent Miclet, Amedeo Napoli

A study of Algorithms Relating Distributive Lattices, Median Graphs, and Formal Concept Analysis

Alain Gély[a,*], Miguel Couceiro[b], Laurent Miclet[c], Amedeo Napoli[b]

[a] *Université de Lorraine, CNRS, LORIA, F-57000 Metz, France*
[b] *Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France*
[c] *Université de Rennes, CNRS, IRISA, Rue de Kérampont, 22300 Lannion, France*

Abstract

In this paper, we study structures such as distributive lattices, distributive semilattices, and median graphs from an algorithmic point of view. Such structures are very useful in classification and phylogeny for representing lineage relationships for example. A distributive lattice can be considered as a median graph while a distributive ∨-semilattice can be considered as a median graph provided that some conditions holding on triple of elements are satisfied. Starting from a lattice structure with different representations, we study the problem of building a median graph from such structures. We make precise and propose algorithms for checking how a lattice can be distributive and can be a median graph. Then, we adapt the problem to semilattices as a lattice where the bottom element is removed is a ∨-semilattice. We also state the problem in terms of Formal Concept Analysis and the representation of a lattice as a formal context, i.e., a binary table. Moreover, we also propose as input a system of implications such as the Duquenne-Guigues basis of a lattice, and we study how to compute such a basis for a distributive semilattice. In the paper, we provide algorithms and examples which illustrate the difficulties related to these different classification tasks. In particular, the minimality of the output lattices is a condition which is hard to ensure and which cannot be always achieved.

1. Motivations

Lattices and median graphs are two structures useful in many applications, in particular in classification and knowledge discovery [1, 2]. Median graphs are especially used in biology, for example in phylogeny, for modeling inter-species relationships [3, 4, 5, 6]. In phylogeny, one of the main problems is to discover evolution trees for representing existing species from accessible DNA

*Corresponding author
Email addresses: `alain.gely@loria.fr` (Alain Gély), `miguel.couceiro@loria.fr` (Miguel Couceiro), `laurent.miclet@gmail.com` (Laurent Miclet), `amedeo.napoli@loria.fr` (Amedeo Napoli)

Preprint submitted to International Journal of Approximate Reasoning *January 20, 2022*

fragments. When several trees are leading to the same inter-species phylogenetic relationships, the preferred ones are the most "parsimonious", where the number of modifications such as mutations for example, is minimal for the considered species. However, several possible parsimonious trees may exist. Such a situation arises with inverse or parallel mutations, e.g., when a gene goes back to a previous state or the same mutation appears for two non-linked species. This calls for a generic representation of such a family of trees.

In Bandelt *et al.* [7, 4], authors propose the notion of *median graph* to deal with this problem, since it was proved that a median graph may encode all parsimonious trees. It is also known that median graphs are related to lattices (see, e.g., [8, 7]) as follows. Any distributive lattice is a median graph while any median graph can be considered as a distributive $\vee$-semilattice such that, for all x, y, z, if the supremum of each pair exists, then the supremum of $\{x, y, z\}$ also exists (it follows that a distributive $\vee$-semilattice is not necessarily a median graph). Dually, a median graph can be considered as a distributive $\wedge$-semilattice such that, for all x, y, z, if the infimum of each pair exists, then the infimum of $\{x, y, z\}$ also exists.

Formal Concept Analysis (FCA) is based on lattice theory and can be used in classification and knowledge discovery [9]. In [10, 11] Uta Priss proposes for the first time to reuse the algorithmic machinery of FCA and the links between distributive lattices and median graphs to analyze phylogenetic trees. However, not every concept lattice is distributive, and thus does not correspond to a median graph. A transformation should be designed to build a median graph from a concept lattice. Moreover, in [11] Uta Priss sketches an algorithm to convert any concept lattice into a median graph. The key step is to transform any concept lattice into a distributive lattice. However, how to transform a concept lattice into a distributive one is not detailed in [11].

An example for relating a median graph and a concept lattice. In [5], Bandelt uses a data set borrowed from [3] to illustrate and evaluate a median graph. In this introduction, we will reuse this example to highlight the commonalities and differences between a median graph and a concept lattice.

The data set is an extract of mitochondrial DNA for 9 groups of individuals from a Khoisan-speaking hunter-gathered population in Southern Africa. For some sequences in mitochondrial DNA (nucleotide positions, denoted by a, b, $\ldots$, j in the data table), a binary information indicates whether a group of individuals owns the consensus version of the sequence –marked as an empty cell in the data table– or a variation for this sequence –marked as a $\times$ in the corresponding cell in the data table. For each nucleotide position, the consensus version of the sequence is the one with the larger number of apparitions in the population. A sequence which is not the consensus version of the sequence is considered as a variation. The data table and the related consensus graph are shown in Table 1 and in Fig. 1 It can be noticed that the data table can fully be considered as a formal context from which a concept lattice can be issued, as it will be discussed below. In this table there are only 8 rows or individual groups (from 0 to 7). The potential ninth row is omitted as it does not present any

	a	b	c	d	e	f	g	h	i	j
0							×			×
1	×				×					
2			×							×
3	×								×	
4							×	×		×
5				×						
6		×						×		
7			×			×				×

Table 1: In this data table borrowed from [5], a row stands for a groups of individual and a column for a variation in the mitocondrial DNA sequence.

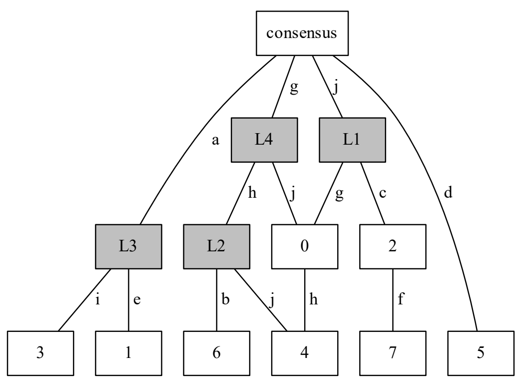

Figure 1: The median graph related to Table 1.

variation on any nucleotide position (consensus group) and thus is not marked with any cross.

In the median graph, vertices stands either for an individual group (numbered from 0 to 7 or *consensus*) or a latent vertex ($L1, L2, L3, L4$). A latent vertex corresponds to an hypothetical group of individuals which is not present in the data set. The existence of such a group is related to the “principle of parsimony”. This principle assumes that, in an evolutionary process, there is no chance for two variations to exactly arise at the same moment in the same population. Thus, only one variation exists between two vertices –groups of individuals in the data set or latent vertices– in the nucleotide sequence. This variation is indicated over the edges. For example, there is only one variation from the consensus group to vertex $L4$ occurring in sequence g while there is again only one variation between vertex $L4$ and vertex 0 occurring in sequence j.

As recalled above, the median graph contains every parsimonious tree as a

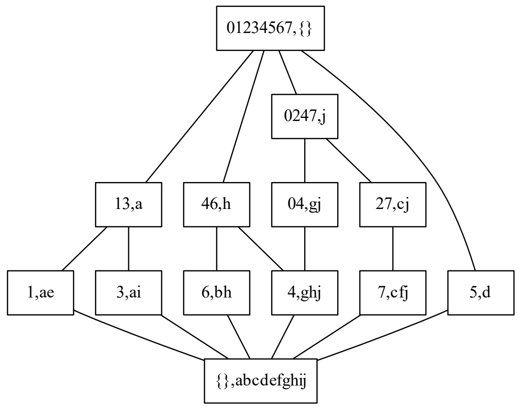

Figure 2: The formal concept lattice issued from the data table displayed in Table 1.

covering tree. A median graph shows some other good properties. Removing an edge labeled with a sequence variation produces two disconnected components, where one component corresponds to individuals holding the variation and the other component to the complementary, i.e., individuals not holding the variation. For example, the edges labeled with j partition the vertices in two connected components. The first component includes the vertices $L1$, 0, 2, 4, and 7, i.e., individuals holding a variant on position j. The complementary component includes all other vertices and the individual groups not holding this variant.

By contrast, the individual groups represented by vertex 0 are related to a double variation denoted by gj. In this case, there is not enough data to determine whether the evolutionary path from the consensus started with a variant on position g, yielding the latent vertex $L4$, or with a variant on j, yielding latent vertex $L1$. Actually, the related median graph is able to display both alternatives, which would not be possible within a parsimonious tree.

In Table 1, the variations in the mitocondrial DNA sequences are represented as a binary data table. Accordingly, FCA algorithms can be applied to build a concept lattice which is shown in Fig. 2. In general, such a concept lattice does not correspond to a median graph. However, as FCA proposes a range of effective and efficient practical tools, the main idea of the present research work is to discuss and to show how to reuse FCA machinery to deal with phylogenetic data and median graphs. Following this idea, the binary data set can be modified in such a way that there exists a bijection between the resulting concept lattice –minus the bottom element– and a median graph. An example of such a transformation is given in Fig. 3. A new column k is added in the formal context (see Table 2) such that a new concept $(046, k)$ appears in the new concept

	a	b	c	d	e	f	g	h	i	j	k
0							×			×	×
1	×				×						
2			×							×	
3	×								×		
4							×	×		×	×
5				×							
6		×						×			×
7			×			×				×	

Table 2: A new column k is added to the data table 1 for building a concept lattice which is a distributive $\vee$-semilattice when $\perp$ is removed.

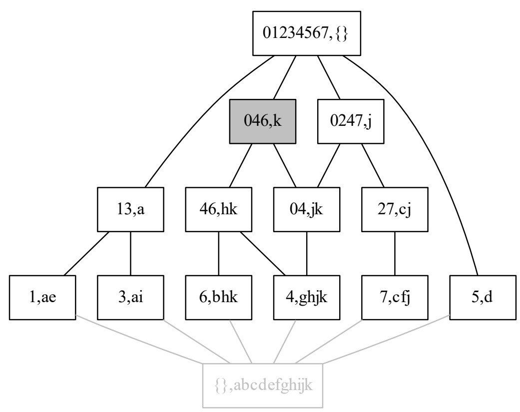

Figure 3: The new concept lattice issued from the modified data table displayed in Table 2 which yields a median graph when the bottom element is omitted.

lattice, which corresponds to $L4$ in the median graph. This concept was missing in the concept lattice computed from the initial data set and displayed in Fig. 2 (a concept lattice which by the way does not correspond to a median graph). It should be noticed that the choice of the new column k is not unique and is guided by several constraints, among which building a concept lattice as close as possible to the initial median graph and minimal in the number of elements to be added.

Outline of the paper. This paper proposes a synthesis and an extension of two preceding papers published in [12, 13], which are based on representation theorem of Birkhoff for distributive lattice [14] and on Formal Concept Analysis [9]. In this paper, we firstly gather some preceding results about the design of distributive lattices from several representations. We make precise the related algorithms and we give examples for illustrating our approach. In particular, we discuss the constraints related to distributivity but also the minimality of the designed structures. It appears that minimality is a hard constraint that cannot be always guaranteed. In addition, we start from several possible representations of lattices and semilattices, and an originality of this paper is to study how the Duquenne-Guigues implication basis, which is minimal and non-redundant, can be used for solving the problem of designing a median graph. This opens the way to an original and very interesting research line in the domain of distributive lattices.

The summary of the paper is as follows. Basic definitions and notations are introduced in Section 2, while Section 3 considers the problem of embedding a lattice or a semilattice into a median graph. In subsection 3.1, we discuss how a lattice can be embedded into a lattice isomorphic to a median graph, while in subsection 3.2, we discuss the same problem for a semilattice. Algorithms are proposed and made precise in Section 4. Firstly, the inputs and outputs of algorithms are lattices or semilattices (see Section 3.1). Secondly, alternative representations such as contexts (see subsection 4.2) or implication bases (see subsection 4.3) are used as input and output of the algorithms.

2. Definitions and Notations

For the sake of completeness, we recall in this section definitions and notations needed throughout the paper. Moreover, we refer the reader to [15, 16] for more details about ordered sets and to [9, 17] for more details about FCA.

2.1. Partially ordered sets and lattices

Only finite sets are considered in this paper. Given a poset $(L, \leq)$ and $X \subseteq L$, $\downarrow X = \{y \in L : y \leq x \text{ for some } x \in X\}$ is the order ideal of X, while $\uparrow X = \{y \in L : x \leq y \text{ for some } x \in X\}$ is the order filter of X. The principal ideal of x is given by $\downarrow x$ (or $\downarrow \{x\}$) and the principal filter by $\uparrow x$ ($\uparrow \{x\}$).

Given a poset $(L, \leq)$ and $x, y \in L$, $x \vee y$ denotes the supremum of x and y while $x \wedge y$ denotes the infimum. A $\vee$-semilattice $(L, \leq)$ is an ordered set such that the supremum exists for all $X \subseteq L$, $X \neq \emptyset$. Dually, a $\wedge$-semilattice $(L, \leq)$ is an ordered set such that the infimum exists for all $X \subseteq L$. A lattice $(L, \leq)$ is an ordered set such that a supremum and an infimum exist for all $X \subseteq L$. Such an order has a lowest element (bottom) denoted by $\bot = \bigwedge L$, and a greatest element (top) denoted by $\top = \bigvee L$.

Given a lattice $(L, \leq)$, an element $x \in L$ such that $x = y \vee z$ implies $x = y$ or $x = z$ is a $\vee$-irreducible element. Dually, an element $x \in L$ such that $x = y \wedge z$ implies $x = y$ or $x = z$ is a $\wedge$-irreducible element. The sets of $\wedge$-irreducible elements and $\vee$-irreducible elements of L are respectively denoted by $\mathcal{M}(L)$ and $\mathcal{J}(L)$. Atoms are minimal $\vee$-irreducible elements while co-atoms are maximal $\wedge$-irreducible elements. $Atoms(L)$ and $Coatoms(L)$ respectively denote the sets of atoms and co-atoms of L.

A lattice L is *distributive* if $\wedge$ and $\vee$ are distributive operations, i.e., for every $x, y, z \in L$, one or equivalently both of the following identities hold:

(i) $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$, $\quad (ii)$ $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

A $\vee$-semilattice SL is distributive if for all $x \in SL$, the lattice defined by $\uparrow x$ is distributive, while a $\wedge$-semilattice SL is distributive if for all $x \in SL$, the lattice defined by $\downarrow x$ is distributive. Since any sublattice of a distributive lattice is distributive, it is sufficient to check distributivity of $\uparrow x$ for x minimal element in SL, or dually distributivity of $\downarrow x$ for x maximal element in SL.

A main result about distributive lattices is given by "representation theorem of Birkhoff" [14], which is based on order ideals. Let $(L, \leq)$ be a poset, $\mathcal{ID}(L) = \{\bigcup_{x \in 2^L} \downarrow X\}$ is the set of order ideals of L. Then, the representation theorem of Birkhoff for distributive lattices states that [14]:

Theorem of Birkhoff. For every poset L, the set $\mathcal{ID}(L)$ ordered by inclusion is a distributive lattice called the *ideal lattice* of L. Moreover, the poset of $\vee$-irreducible elements of $\mathcal{ID}(L)$ given by $\mathcal{J}(\mathcal{ID}(L)) = \{\downarrow x \mid x \in L\}$ is order-isomorphic to L.

2.2. Median Graphs

The following property discussed in [4] establishes links existing between median graphs and distributive lattices.

Definition and properties of a median graph. A graph is a median graph if and only if it is isomorphic to a $\vee$-semilattice L with the two following properties:

i) DIST: L is distributive,
ii) TRIPLE: for all $x, y, z \in L$ such that $(x \wedge y)$, $(y \wedge z)$, and $(z \wedge x)$ are defined, then $(x \wedge y \wedge z)$ is defined.

Thanks to duality between $\vee$ and $\wedge$, the corresponding property exists in a distributive $\wedge$-semilattice.

In the following and for the sake of simplicity, we will use the expressions "median lattice" and "median semilattice" for denoting a distributive lattice (respectively semilattice) verifying the conditions DIST and TRIPLE.

2.3. Formal Concept Analysis

Formal Concept Analysis [9, 17] is a mathematical formalism based on lattice theory and aimed at data analysis and knowledge discovery. FCA allows to build a concept lattice from a formal context (G, M, I) using two derivation operators. In (G, M, I), G is a set of objects, M a set of attributes, and I an incidence relation between objects and attributes. Given $X \subseteq G$ and $Y \subseteq M$, the derivation operators denoted by $'$ work as follows:

$X' = \{y \in M \mid xIy \text{ for all } x \in X\}$ is the set of attributes common to objects in X while $Y' = \{x \in G \mid xIy \text{ for all } y \in Y\}$ is the set of objects having all attributes in Y.

A formal concept is a pair (X, Y) verifying $X' = Y$ and $Y' = X$. The extent X and the intent Y of the concept (X, Y) are closed sets verifying $X = X''$ and

$Y = Y''$ (derivation can be iterated). The set of all formal concepts ordered by inclusion of extents generates a concept lattice. Moreover, the set of extents and of intents are closure systems, i.e., a family of sets closed by intersection with a maximal element. The operator $''$, which is the composition of the two derivation operators is a closure operator (a closure operator is monotone, extensive, and idempotent).

To terminate, we recall basics about implications and implication bases in concept lattices. Given a context (G, M, I), an implication between attributes in M is a pair of subsets of M, denoted by $A \rightarrow B$. A subset $T \subseteq M$ respects the implication $A \rightarrow B$ if $A \not\subseteq T$ or $B \subseteq T$.

Farther, we will use the notion of minimum basis of implications, i.e., a family of implications Σ which is minimal among all equivalent families of implications (they represent the same set of implications). In FCA, the so-called canonical basis or Duquenne-Guigues basis shows very interesting properties [18].

Let $X \rightarrow X''$ be a closure operator on M. A subset $P \subseteq M$ is pseudo-closed iff (i) $P \neq P''$, and (ii) if $Q \subset P$ is a pseudo-closed proper subset of P, then $Q'' \subseteq P$.

Then the canonical basis of Duquenne-Guigues can be defined as the set of implications $\{P \rightarrow P'' \,|\, P \text{ is pseudo-closed}\}$. This basis is sound, complete w.r.t. the operator $''$, and non-redundant.

3. Embedding a Semilattice Into a Median Graph

In this section, firstly we investigate the following problem: considering a lattice L given as input, how to embed L into a median graph, in fulfilling some properties about minimality or maximality w.r.t. the size of the median graph. Secondly, we consider a variation of the above problem. While no particular information is usually associated with the element $\perp$ in a lattice, given as input a $\vee$-semilattice SL –then $SL \cup \{\perp\}$ is a lattice– how to embed SL into a median graph.

3.1. Embedding a Lattice Into a Lattice Isomorphic to a Median Graph

A median graph is equivalent to a distributive $\vee$-semilattice SL when the TRIPLE condition is met. It should be noticed that if SL is a lattice, then the TRIPLE condition is obviously satisfied since the infimum exists for every subset of elements in SL. Thus, firstly we check how to embed a lattice L into a median graph which is also a lattice and not only a $\vee$-semilattice.

First order embedding proposal.

The problem can be formally stated as follows:

Problem 1. Basic order-embedding
Input: a lattice L,
Output: a lattice LM such that:

i) L can be order-embedded into LM,

ii) LM is a median graph.

Since LM is both a median graph and a lattice, then it follows that LM is a distributive lattice. It is straightforward that L can be order-embedded into a Boolean lattice, i.e., a distributive lattice isomorphic to $(2^{\mathcal{J}(L)}, \subseteq)$. Actually, this output is not really interesting in data analysis tasks. Indeed, two different lattices L_1 and L_2 having the same number of $\vee$-irreducible elements, i.e., $|\mathcal{J}(L_1)| = |\mathcal{J}(L_2)|$, will be mapped into the same Boolean lattice whereas they can be very different.

Second order-embedding proposal with invariance of $(\mathcal{J}(L), \leq)$.

Regarding order-embedding, the Boolean lattice $(2^{\mathcal{J}(L)}, \subseteq)$ is a maximal output w.r.t. the number of elements, but this embedding does not yield meaningful information. Recall that the posets of irreducible elements, namely $(\mathcal{J}(L), \leq)$ and $(\mathcal{M}(L), \leq)$, are important suborders of a lattice as they can be used to rebuild the whole lattice L. Moreover, we keep in mind that two lattices LM_1 and LM_2 may differ despite $|\mathcal{J}(L_1)| = |\mathcal{J}(L_2)|$. Any two lattices L_1 and L_2 are considered to be isomorphic iff $(\mathcal{J}(L_1), \leq)$ is isomorphic to $(\mathcal{J}(L_2), \leq)$ and $(\mathcal{M}(L_1), \leq)$ is isomorphic to $(\mathcal{M}(L_2), \leq)$. Then there does not exist any isomorphism between L_1 and L_2 when one or both posets differ, i.e., $(\mathcal{J}(L_1), \leq) \neq (\mathcal{J}(L_2), \leq)$ and/or $(\mathcal{M}(L_1), \leq) \neq (\mathcal{M}(L_2), \leq)$.

To enforce the output LM to be more closely related to the input lattice L, the following constraint relating L and LM can be considered: either $(\mathcal{J}(LM), \leq)$ is isomorphic to $(\mathcal{J}(L), \leq)$ or $(\mathcal{M}(LM), \leq)$ is isomorphic to $(\mathcal{M}(L), \leq)$. Here after, we assume that $(\mathcal{J}(LM), \leq)$ is isomorphic to $(\mathcal{J}(L), \leq)$ and we investigate the following problem:

Problem 2. Embedding with invariance of $(\mathcal{J}(L), \leq)$
Input: a lattice L,
Output: a lattice LM such that:

i) L can be order-embedded into LM,
ii) LM is a median graph,
iii) $(\mathcal{J}(L), \leq) = (\mathcal{J}(LM), \leq)$.

A solution is directly given by the representation theorem of Birkhoff [14], where it is shown that there exists an isomorphism between the ideal lattice and a distributive lattice. It follows that the ideal lattice of $(\mathcal{J}(L), \leq)$ is a distributive lattice such that L can be order-embedded into $(\mathcal{J}(L), \leq)$. In this case, there is a $\wedge$-embedding from L into LM, and LM is a maximal lattice with respect to the embedding with $(\mathcal{J}(L), \leq) = (\mathcal{J}(LM), \leq)$ [19].

Moreover, if we have $LM \neq LD$ where LD is a distributive lattice such that LD can be order-embedded into LM, then $(\mathcal{J}(LM), \leq)$ is not isomorphic to $(\mathcal{J}(LD), \leq)$, and thus the ideal lattice of $(\mathcal{J}(L), \leq)$ is the greatest lattice with respect to inclusion verifying $(\mathcal{J}(L), \leq) = (\mathcal{J}(LM), \leq)$.

3.2. Embedding a Semilattice Into a Semilattice Isomorphic to a Median Graph

In the previous subsection, we investigate the embedding of a lattice into a lattice isomorphic to a median graph. Here we do not consider anymore a lattice as input but instead a $\vee$-semilattice. This is justified by the fact that a median graph is usually not a lattice. More precisely we study the following problem:

Problem 3. Embedding with invariance of $(\mathcal{J}(SL), \leq)$
Input: a $\vee$-semilattice SL,
Output: SM a $\vee$-semilattice such that:

i) SL can be order-embedded into SM,
ii) SM is a median graph,
iii) $(\mathcal{J}(SL), \leq) = (\mathcal{J}(SM), \leq)$.

As all sets are assumed to be finite, it should be noticed that a $\vee$-semilattice SL always verifies that $\{SL \cup \bot\}$ is a lattice. In this way, given a $\vee$-semilattice SL, $(\mathcal{J}(SL), \leq)$ and $(\mathcal{M}(SL), \leq)$ denote the set of irreducible elements of the lattice $\{SL \cup \bot\}$. Then, one could argue that it could be sufficient to reuse and apply the preceding transformation to the semilattice $\{SL \cup \bot\}$ and build a distributive lattice say SLM having the same set of $\vee$-irreducible elements. This is possible but not optimal as shown here after.

If a lattice L is distributive, then $L \backslash \bot$ is a distributive $\vee$-semilattice. Next, a $\vee$-semilattice SL is distributive iff for each element $e \in SL$, $\uparrow e$ is a distributive lattice. $\uparrow e$ is a sublattice of SL and a sublattice of a distributive lattice is distributive. More than that, a $\vee$-semilattice SL is distributive iff for all *minimal* element $e \in SL$, $\uparrow e$ is a distributive lattice. The converse is not true in general: $L \backslash \bot$ may be a distributive $\vee$-semilattice and L be a nondistributive lattice.

Two simple examples are the N_5 and M_3 lattices. Recall that a lattice L is not distributive iff the diamond M_3 or the pentagon N_5 is order-embedded into L (see for example [15]). An embedding of the lattice M_3 –more generally M_n where n is the number of atoms– into a distributive lattice with the same poset of $\vee$-irreducible element yields a Boolean lattice isomorphic to $(2^n, \subseteq)$. Thus a lattice with $n+2$ elements such as M_n can be embedded into a lattice with 2^n elements, i.e., $(2^n, \subseteq)$.

Now, consider the $\vee$-semilattice $SM_n = M_n \backslash \bot$, where the atoms of M_n are the minimal elements of SM_n. Given an atom $a \in M_n$, the filter $\uparrow a = \{a, \top\}$ determines a total order and hence a distributive lattice. Then the lattice M_n can be embedded into a Boolean lattice with 2^n elements while the semilattice SM_n is isomorphic to a median graph and can be embedded into itself. This shows that the embedding of a $\vee$-semilattice such as $L \backslash \bot$ can be much more compact w.r.t. the number of elements than the embedding of the related lattice L.

Existence of an infimum in the TRIPLE *condition.*

Recall that a median graph requires the two conditions DIST and TRIPLE to be satisfied. The first condition DIST is related to distributivity while the

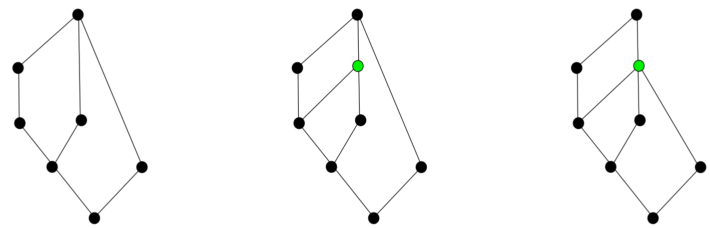

Figure 4: The lattice in subfigure (a) is such that there exist two non isomorphic minimal distributive $\vee$-semi-lattices (b) and (c) when the bottom element is removed.

second condition TRIPLE involves the existence of particular infimum elements. Previously, as either the input or the output are lattices, the TRIPLE condition is satisfied because of the existence of an infimum for any subset of elements. Now, if a $\vee$-semilattice SL is considered, we must check that the TRIPLE condition is effectively satisfied. As $\{SL \cup \bot\}$ is a lattice, the TRIPLE condition may be rewritten in such a way that it involves the $\bot$ element:

$$\text{if } x \wedge y, y \wedge z, x \wedge z \text{ exist, then } x \wedge y \wedge z \text{ also exists} \qquad (\text{TRIPLE})$$

$$\text{if } x \wedge y > \bot, y \wedge z > \bot, x \wedge z > \bot \text{ then } x \wedge y \wedge z > \bot, \qquad (\text{TRIPLE}_\bot)$$

In Section 4, we make precise the design of algorithms based on the condition $\text{TRIPLE}_\bot$ to compute a median graph as a semilattice or a lattice.

3.3. Minimal and minimum embedding

Given a semilattice SL, a subsequent question is to check whether there exists a minimum, i.e., minimal and unique, $\vee$-semilattice SM such that SL can be embedded into SM through an isomorphism between posets of $\vee$-irreducible elements. Actually, the answer is negative since a counter-example showing that such a minimum does not always exist was proposed in [20] and is displayed in Fig. 4.

Given the lattice in subfigure (a), the lattices in subfigures (b) and (c) differ by only one element and are minimal distributive $\vee$-semilattices when the $\bot$ element is removed. However, (b) and (c) are not isomorphic and no minimum $\vee$-semilattice exists. Then a practical strategy should be defined when using FCA algorithms to build median graphs. In the following section we show that the proposed algorithms do not always output a minimal solution.

4. Three Algorithms for Computing Distributive Lattices and Median Graphs

4.1. An Algorithm Based on a Lattice Structure

In this section, we propose an algorithm which takes a $\vee$-semilattice as input and returns as output a median graph. Below we explain the process and we prove the correctness of the algorithm. Then we adapt the algorithm to deal with alternate representations of a concept lattice such as a context or a basis of implications.

Algorithm 1: Construction of a median ($\vee$-semi) lattice.

Data: A semilattice SL
Result: a semilattice SM and a Boolean $\text{TRIPLE}_\perp$
$\text{TRIPLE}_\perp \leftarrow check\_\text{TRIPLE}_\perp\_condition(SL)$
if $\text{TRIPLE}_\perp == false$ **then**
  return $\mathcal{ID}(\mathcal{J}(SL \cup \perp)), \subseteq), \text{TRIPLE}_\perp$
$SM \leftarrow SL$; **repeat**
  stability $\leftarrow$ true;
  foreach $a \in atoms(SM)$ **do**
    compute P_a the poset of $\vee$-irreducible elements in $\uparrow a$
    $SM \leftarrow SM \cup \mathcal{ID}(P_a)$ (embed the lattice $\uparrow a$ into the ideal lattice of $\vee$-irreducible elements)
    if *SM modified since last iteration* **then**
      stability $\leftarrow$ false;
until *stability*
return $\{(SM, \subseteq), \text{TRIPLE}_\perp\}$

The algorithm implements the two conditions for a structure to be a median graph, i.e., checking whether the output satisfies the $\text{TRIPLE}_\perp$ condition and whether every atom filter in the output is a distributive lattice.

In algorithm 1, the $\text{TRIPLE}_\perp$ condition is verified in polynomial time w.r.t. the size of the lattice (checking for every triple). If $\text{TRIPLE}_\perp$ is not satisfied, then the algorithm outputs the ideal lattice of $(\mathcal{J}(SL), \leq)$. This is the only solution where $(\mathcal{J}(SL), \leq)$ is invariant. When $\text{TRIPLE}_\perp$ is satisfied, then the algorithm computes a $\vee$-semilattice, which can be considered as a lattice when the element $\perp$ is added. The idea is to embed lattices defined by the filter of atoms into a distributive lattice using the representation theorem of Birkhoff.

It should be noticed that the $\text{TRIPLE}_\perp$ condition is checked over the input lattice, thanks to propositions 1 and 2 which are presented here after. Proposition 1 states that if the input concept lattice does not satisfy the condition $\text{TRIPLE}_\perp$, the output cannot satisfy $\text{TRIPLE}_\perp$ either. Proposition 2 states that if the input satisfies the condition $\text{TRIPLE}_\perp$, then the output of the algorithm should satisfy $\text{TRIPLE}_\perp$.

Lemma 1. *Let L and LD be two lattices such that L can be embedded into LD and $(\mathcal{J}(L), \leq)$ is isomorphic to $(\mathcal{J}(LD), \leq)$. If $\text{TRIPLE}_\perp$ is not satisfied in L, then $\text{TRIPLE}_\perp$ cannot be satisfied in LD.*

Proof. Since L and LD are two lattices such that L can be embedded into LD and $(\mathcal{J}(L), \leq)$ is isomorphic to $(\mathcal{J}(LD), \leq)$, the embedding from L to LD is $\wedge$-preserving. It follows that if there exist $x, y, z \in L$ such that $x \wedge y > \perp$,

$x \wedge z > \bot$, $y \wedge z > \bot$, and $x \wedge y \wedge z = \bot$, then there exist three elements $f(x)$, $f(y)$, and $f(z)$ in LD, images of x, y, z through the embedding, such that $f(x) \wedge f(y) > \bot$, $f(x) \wedge f(z) > \bot$, $f(y) \wedge f(z) > \bot$, and $f(x) \wedge f(y) \wedge f(z) = \bot$, and thus TRIPLE$_\bot$ is not satisfied in LD.

The following proposition is a consequence of the lemma:

Proposition 1. *If L is a lattice in which there exist $x, y, z \in L$ with $x \wedge y, x \wedge z, y \wedge z > \bot$ and $x \wedge y \wedge z = \bot$, then there is exactly one median semilattice S_L such that $(\mathcal{J}(L), \leq_L)$ is isomorphic to $(\mathcal{J}(S_L), \leq_{S_L})$. This semilattice is the ideal lattice of $(\mathcal{J}(L), \leq_L)$.*

PROOF. Let us consider a lattice L and a semilattice S_L verifying the hypotheses. Let us add the new element $a = x \wedge y \wedge z$ to S_L and check whether S_L is a median semilattice. There are two alternatives which are illustrated in Fig. 5.

1. $a = \bot$ and in this case S_L is a lattice. In addition, S_L is a median semilattice if it is distributive, and thus corresponds to the ideal lattice of $(\mathcal{J}(L), \leq)$.
2. $a \neq \bot$ and in this case a is a new atom in S_L and thus a new $\vee$-irreducible element. Then it can be checked that there is no isomorphism between $(\mathcal{J}(L), \leq)$ and $(\mathcal{J}(L_d), \leq)$

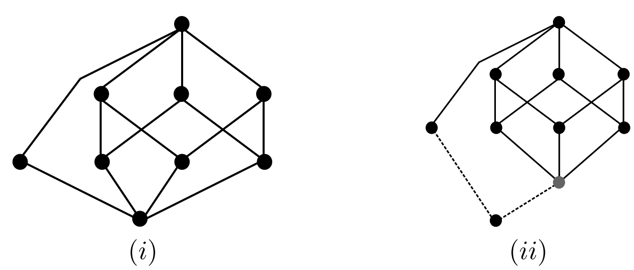

Figure 5: Illustration of the proof of Proposition 1. In the lattice (i), each filter is distributive and $L \backslash \bot$ is a distributive $\vee$-semilattice. However, TRIPLE$_\bot$ is not satisfied. A first alternative is to keep $\bot$ in L but then the lattice (i) should be distributive and this is not the case. Another alternative is to add a new infimum distinct from $\bot$. Then the posets of $\vee$-irreducible elements of (i) and (ii) are no more isomorphic, which violates the hypothesis of the proposition.

When the condition TRIPLE$_\bot$ is satisfied, it remains to ensure that every filter of $L \backslash \bot$ is distributive. This process is performed as follows. The filter of an atom a is considered as an independent lattice denoted by L_a and the theorem of Birkhoff is used to compute the ideal lattice of $(\mathcal{J}(L_a))$. The ideal lattices of the atoms are combined using the ideal representation of each atom and the inclusion order. At this step, it may be possible that an element in the filter of the atom a belongs to the filter of another atom b and breaks the distributivity in $\uparrow b$. In this case, the process of adding a new element is repeated until a fixpoint is reached, i.e., it is no more necessary to add a new element as distributivity is achieved. Such a fixpoint exists, since at each step, only new elements are added and at most the ideal lattice of $(\mathcal{J}(L), \leq)$ is obtained (See Fig. 6). Moreover, from one step to the other, the condition $\text{TRIPLE}_{\perp}$ remains satisfied.

Proposition 2. *Let us consider Algorithm 1 and assume that SM_i is the lattice used in iteration i. If $\text{TRIPLE}_{\perp}$ is satisfied in iteration i then $\text{TRIPLE}_{\perp}$ is satisfied in iteration $i+1$,*

Proof. In each iteration i of algorithm 1, i.e., an iteration of the "repeat" loop, a lattice SM_i is embedded into a new lattice SM_{i+1} such that $(\mathcal{J}(SM_i), \leq)$ is isomorphic to $(\mathcal{J}(SM_{i+1}), \leq)$. Let us assume that:

1. in SM_i, the condition $\text{TRIPLE}_{\perp}$ is satisfied,
2. in SM_{i+1}, the condition $\text{TRIPLE}_{\perp}$ is not satisfied, i.e., there exist e_1, e_2, e_3 such that $e_1 \wedge e_2 \wedge e_3 = \perp$, $e_1 \wedge e_2 = x \neq \perp$, $e_1 \wedge e_3 = y \neq \perp$, and $e_2 \wedge e_3 = z \neq \perp$.

We will show that these two hypotheses are leading to a contradiction. When $\text{TRIPLE}_{\perp}$ is satisfied in SM_i, at least one of the three elements e_1, e_2, e_3 should not belong to SM_i, say e_3 (without loss of generality). Then e_3 can be considered as a new element where $y < e_1$ and $y < e_3$ since $y = e_1 \wedge e_3$. As well $z < e_2$ and $z < e_3$ as $z = e_2 \wedge e_3$.

In the algorithm, the new element e_3 is introduced in SM_{i+1} because the filter of an atom a is not distributive in SM_i. Thus there exists an atom a with $a < y < e_3$ and $a < z < e_3$. Moreover, $a < y < e_1$ and $a < z < e_2$. To sum up, it comes that $a < e_1$, $a < e_2$, $a < e_3$ and then it follows that $e_1 \wedge e_2 \wedge e_3 \geq a > \perp$. This leads to a contradiction with hypothesis 2 above.

Thus we may conclude that if the condition $\text{TRIPLE}_{\perp}$ is satisfied at the beginning of an iteration i, then it is satisfied at the beginning of the next iteration $i+1$.

In Algorithm 1, the filters of the atoms are considered independently one from the others. This approach is correct but does not necessarily output a minimal embedding, i.e., there may exist a distributive $\vee$-semilattice SM_m that can be embedded in SM, while SL can be embedded into SM_m with $|SD| < |SM_m| < |SM|$ and $(\mathcal{J}(SL), \leq) = (\mathcal{J}(SM_m), \leq) = (\mathcal{J}(SM), \leq)$.

This may happen because each filter is independently processed, and an element may belong to several filters, then processing one filter may break the distributivity of another one. This is illustrated in Fig. 6. Element g is added in (ii) to make distributive the left atom filter –denoted by 1 in (ii). However, applying the same process for making distributive the filter of the rightmost atom –denoted by 4 in (ii)– the filter of 1 is no more distributive. A local modification of the filters, e.g., adding the element r in the filter of 4 but also in the filter of 1, may have an impact on neighboring filters and breaks the distributivity. The processing of filters consists in adding elements until no

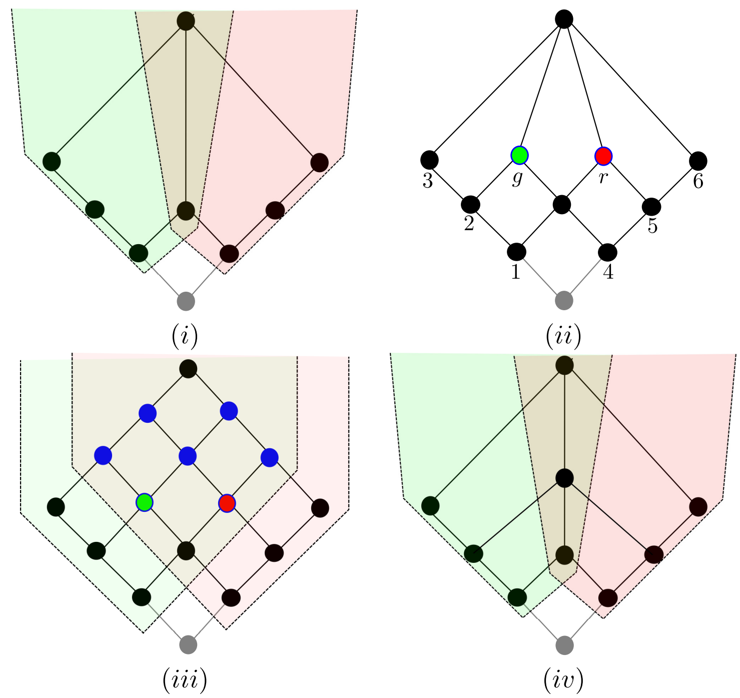

Figure 6: *From left to right and top to bottom:* (i) The input lattice: one element is at the intersection of two filters. (ii) During the first iteration in the loop of Algorithm 1, g is added for making the filter $\uparrow 1$ distributive. Similarly, r is added for making the filter $\uparrow 4$ distributive. However, g belongs to $\uparrow 4$ while r belongs to $\uparrow 1$, which breaks distributivity in $\uparrow 1$ and $\uparrow 4$ at the end of the first iteration. Thus new elements should be added. (iii) After a finite number of iterations, no more elements need to be added to the filters as they are all distributive. The output of the algorithm is a lattice structure. (iv) The semilattice $(i)\backslash\bot$ can be embedded into the semilattice $(iv)\backslash\bot$ which is a median graph with the same poset of $\vee$-irreducible elements as lattice (i).

more change arises and distributivity is satisfied. The resulting structure is a median graph, i.e., a distributive $\vee$-semilattice satisfying the TRIPLE condition. However, there is no guarantee that this resulting structure is of minimal size w.r.t. the number of elements. In particular, it should be noticed that lattice (iv) in Fig. 6 is a minimal lattice verifying all the constraints and where only one element is added w.r.t. the initial lattice (i). However, this lattice cannot be designed by the algorithm, showing that minimality of the output lattice cannot be guaranteed.

4.2. An Algorithm Based on a Formal Context

In this subsection, we present an algorithm whose input is a standard context –as used in FCA or in phylogeny– and whose output is the context related to a distributive lattice verifying the properties of a median graph, namely DIST and TRIPLE.

Firstly, let us recall that the standard context of the ideal lattice of a poset P is given by $(P, P, \not\geq)$ [9]. The ideal lattice of $\vee$-irreducible elements $(J(P), \leq)$ can be computed from the context $(P, P, \not\geq)$. Moreover, embedding every atom filter into a distributive lattice can be performed in polynomial time in the size of the context. Then it remains to check the TRIPLE condition when the input is a context. To sum up, we present in the following the design of an algorithm deciding in polynomial time in the size of a context, i.e., $\mathcal{J}(L) \times \mathcal{M}(L)$, whether the condition TRIPLE is satisfied. This is the objective of the following lemma.

Lemma 2. *Let L be a lattice. There exist $x, y, z \in L$ such that $x \wedge y, x \wedge z, y \wedge z > \bot$ and $x \wedge y \wedge z = \bot$* iff *there exist $a_x, a_y, a_z \in Atoms(L)$, and $m_x, m_y, m_z \in \mathcal{M}(L)$, such that:*

- $a_x \not\leq m_x$, $a_x \leq m_y$, $a_x \leq m_z$,
- $a_y \not\leq m_y$, $a_y \leq m_x$, $a_y \leq m_z$,
- $a_z \not\leq m_z$, $a_z \leq m_x$, $a_z \leq m_y$.

Proof. Suppose that there exist $x, y, z \in L$ such that $x \wedge y, x \wedge z, y \wedge z > \bot$ and $x \wedge y \wedge z = \bot$. Then there exist 3 atoms a_x, a_y, a_z such that: (i) $a_z \leq x \wedge y$ and $a_z \not\leq z$ otherwise $x \wedge y \wedge z \geq a_z > \bot$. Similarly, it comes (ii) $a_y \leq x \wedge z$, $a_y \not\leq y$ otherwise $x \wedge y \wedge z \geq a_y > \bot$, and (iii) $a_x \leq y \wedge z$, $a_x \not\leq x$ otherwise $x \wedge y \wedge z \geq a_x > \bot$.

Now let us pick $m_i \in Max(\mathcal{M}(L) \setminus \uparrow a_i)$ with $i \in \{x, y, z\}$. Then, (i) m_x is greater than a_y and a_z and not greater than a_x, and similarly, (ii) m_y is greater than a_x and a_z and not greater than a_y, and (iii) m_z is greater than a_x and a_y and not greater than a_z, showing that the conditions are necessary.

Now, for showing that the conditions are also sufficient, let us suppose that there are $a_1, a_2, a_3 \in Atoms(L)$ and $m_1, m_2, m_3 \in \mathcal{M}(L)$ such that: (i) $a_1 \not\leq m_1$, $a_1 \leq m_2$, $a_1 \leq m_3$, (ii) $a_2 \not\leq m_2$, $a_2 \leq m_1$, $a_2 \leq m_3$, and (iii) $a_3 \not\leq m_3$, $a_3 \leq m_1$, $a_3 \leq m_1$.

A solution is given by $x = a_2 \vee a_3$, $y = a_1 \vee a_3$, and $z = a_1 \vee a_2$. It should be noticed that there cannot be any comparable pair from $\{x, y, z\}$ as this would contradict the existence of $m_1, m_2, m_3 \in \mathcal{M}(L)$ satisfying the conditions of the lemma. Hence, x, y and z are pairwise incomparable w.r.t. $\leq$. Moreover, $x \wedge y > \bot$, $x \wedge z > \bot$, $y \wedge z > \bot$ and $x \wedge y \wedge z = \bot$. This completes the proof of the lemma.

This lemma ensures that it is possible to check whether the TRIPLE$_\bot$ condition is satisfied in polynomial time in the standard context of a lattice. It should be noticed that this problem can be seen as a variation applied to atoms, i.e., minimal $\vee$-irreducible elements, of a result obtained on totally balanced matrices, see for example [21, 22].

Algorithm 2 is based on Lemma 2. The input of the algorithm is a context CT which is a representation of a semilattice, i.e., the concept lattice related to CT minus the element $\bot$. The output of the algorithm is a context and the

Boolean value $\text{TRIPLE}_\perp$. If $\text{TRIPLE}_\perp$ is true, then the output context is the representation of a median semilattice (the lattice without the bottom element $\perp$). If $\text{TRIPLE}_\perp$ is false, then the output context is the representation of the ideal lattice of $(\mathcal{J}(L), \leq)$.

Algorithm 2: Construction of the context of a median ($\vee$-semi) lattice.

Data: A context $CT(L) = (\mathcal{J}(L), \mathcal{M}(L), I)$ of a lattice L.
Result: A context $CT(L_{med}) = (\mathcal{J}(L_{med}), \mathcal{M}(L_{med}), I)$
and the Boolean value $\text{TRIPLE}_\perp$.
$tri \leftarrow check\_\text{TRIPLE}_\perp\_condition((\mathcal{J}(L), \mathcal{M}(L), I))$
if $tri == false$ **then**
  return $CT(\mathcal{J}(L), \mathcal{J}(L), \ngeq), false$
foreach $j \in \mathcal{J}(L)$*, minimal* **do**
  $(P_j, \leq) \leftarrow \emptyset$
repeat
  stability $\leftarrow$ true;
  foreach $j \in \mathcal{J}(L)$*, minimal* **do**
    compute P_j the poset of $\vee$-irreducible elements in $\uparrow j$
    compute $CT_j = (P_j, P_j, \ngeq)$
    if P_j *modified since last iteration* **then**
      stability $\leftarrow$ false;
  Merge all $CT_j = (P_j, P_j, \ngeq)$ in a unique context CT
  Reduce CT
until *stability*
return $CT(L_{med}), true$

4.3. An Algorithm Based on an Implication Basis

In this section, we consider an alternative representation in the FCA formalism which is based on an implication basis. In the preceding section we were interested in building a median semilattice based on a lattice L and on the reduced context (G, M, I) related to L. Here we still study how to build a semilattice SL but this time based on the Duquenne-Guigues basis –DG-basis– of implications, also called the canonical basis of implications (see [9, 17] for example). The objective is to build the DG-basis of a distributive $\wedge$-semilattice starting from the DG-basis of an arbitrary lattice, taking into account $\wedge$-irreducible elements and the fact that any two elements should have a meet.

More practically, for working with implications, we consider a ground set on which are based all implications and which corresponds to J, i.e. the set of $\vee$-irreducible elements, represented by the $\top$ element in the lattice L. A correspondence can be made with FCA and a context (G, M, I), where the ground set would be M. Moreover, for an element m of the lattice L, we denote the "closure" m'' of m as $m'' = \{\downarrow m\} \cap J$.

The main principle on which are relying the algorithms presented above is to decompose an initial lattice into sublattices related to atoms. Each sublattice corresponds to the filter $\{\uparrow a\}$ of an atom a and is modified in order to meet constraints related to a median semilattice (i.e., DIST and TRIPLE conditions). Then starting with the reduced context of a lattice L, one can build the context corresponding to the filter $\{\uparrow a\}$ of an atom a. Such an operation is not so straightforward when the input is the DG-basis of the underlying lattice. There does not exist any easy way to build the DG-basis of the filter $\{\uparrow a\}$ of an atom a or of any other element of the initial lattice L. By contrast, given a co-atom $m \in L$, the DG-basis of the ideal $\{\downarrow m\}$ can be easily built from the DG-basis of L.

This can be illustrated as follows. Let us consider $\Sigma_L = \Sigma_m \cup \Sigma_{\overline{m}}$ the DG-basis of L. Then the DG-basis of $\{\downarrow m\}$ is given by Σ_m, where $m'' = \{\downarrow m\} \cap J$:

- $\Sigma_m = \{P \to P''\}$ with $P'' \subseteq m''$,
- $\Sigma_{\overline{m}} = \{P \to P''\}$ with $P'' \not\subseteq m''$.

Actually, the implications lying in $\Sigma_{\overline{m}}$ are related to elements which are not in m'' and thus which are neither related to any closed set in m'' nor in $\{\downarrow m\}$. Moreover, Σ_m is an implication basis of $\{\downarrow m\}$ where implications are related to closed sets which are lower than m for the lattice order (recall that $\{\downarrow m\}$ is an ideal). In addition, Σ_m is a DG-basis. Let us suppose the contrary, then P is not a pseudo-closed set in at least one implication $P \to P''$ in Σ_m. In such a case, P would not be either a pseudo-closed set in Σ_L which contradicts the initial hypothesis.

Building a distributive $\wedge$-semilattice.

As discussed above, a median graph can be seen as a $\vee$-semilattice or a $\wedge$-semilattice verifying the conditions of distributivity (DIST) and existence of a lower bound (TRIPLE). The algorithms defined above are based on filters in a $\vee$-semilattice, i.e., $L \setminus \bot$. Let us now consider the dual situation and a $\wedge$-semilattice based on ideals. We have just seen above how to build the DG-basis related to the lattice induced by an ideal $\{\downarrow m\}$. Then contrasting what was done in the preceding sections, we will now consider a $\wedge$-semilattice and $L \setminus \top$ instead of $L \setminus \bot$. While the latter condition is changed, we are still trying to build a median semilattice SL (i.e., $L \setminus \top$) where the set of $\vee$-irreducible elements namely J is invariant, and such that the initial lattice L can be embedded into SL. This can be summarized in the following theorem:

Theorem 1. *Let us consider a lattice L and the related DG-basis Σ_L. Then $L \setminus \top$ is a distributive $\wedge$-semilattice iff for every implication $P \to P'' \in \Sigma_L$, $|P| = 1$ or $P'' = \top$.*

PROOF. For proving this theorem, we should recall that the DG-basis of a distributive lattice L is such that [23, 24]: $\Sigma_L = \{j \to j'' \mid j \in (\mathcal{J}(L), \leq)\}$, where $(\mathcal{J}(L), \leq)$ is the set of $\vee$-irreducible elements in L.

Let m be a co-atom in L and suppose that for all $P \to P'' \in \Sigma_L$, either $|P| = 1$ or $P'' = \top$. Intuitively, the condition $P'' = \top$ corresponds to the case

$P'' = J$, i.e., there does not exist a closed element including P without including all other elements.

Then, $\{\downarrow m\}$ is a lattice such that $\Sigma_{\{\downarrow m\}} = \{P \to P'' \mid P \to P'' \in \Sigma_L, P'' \subseteq \{\downarrow m\}\}$. As m is a co-atom, $m < \top$ and $P'' \neq \top$. It follows that every implication $P \to P''$ in $\Sigma_{\{\downarrow m\}}$ is such that $|P| = 1$ and that the DG-basis of the lattice $\{\downarrow m\}$ is of the form $\{j \to j''\}$, i.e., the implication basis of a distributive lattice. Thus $L\backslash\top$ is a distributive $\wedge$-semilattice.

Now, let us assume that $L\backslash\top$ is a distributive $\wedge$-semilattice. Then every ideal is a distributive lattice where the DG-basis is such that $\Sigma_L = \{P \to P''\}$ with $|P| = 1$. Σ_L includes every implication of the form $\{j \to j''\}$ and as well implications $\{P \to P''\}$ such that $P\backslash m'' \neq \emptyset$ for a co-atom $m \in L$. In such a case, P'' can only be $\top$.

This theorem can be implemented within an algorithm designing the DG-basis of a $\wedge$-semilattice $L \setminus \top$ based on the DG-basis of an arbitrary lattice L. If Σ_L denotes the DG-basis of the lattice L, then one can build $\Sigma_{L\setminus\top}$ corresponding to the $\wedge$-semilattice $L \setminus \top$, i.e., $\Sigma_{L\setminus\top} = \{P \to P'' \mid P \to P'' \in \Sigma_L, |P| = 1$ or $P'' = \top\}$.

As $\Sigma_{L\setminus\top} \subseteq \Sigma_L$, then L can be embedded into $L \setminus \top$. Moreover, since $\Sigma_{L\setminus\top}$ and Σ_L have the same set of implications, i.e., $P \to P''$ with $|P| = 1$, the lattices L and $L \setminus \top$ have the same set of $\vee$-irreducible elements.

Finally, for building a median graph, the TRIPLE condition remains to be satisfied based on the DG-basis. This last point should still be investigated in future work.

5. Conclusions and Perspectives

In this paper, we investigated how to build a median graph from a lattice and a semilattice. In particular, we also studied how this can be adapted to the special case of concept lattices in Formal Concept Analysis. We have made precise three algorithms whose output is a median graph. The first one has as input a lattice or a semilattice. The second and third algorithms are working within the formalism of FCA and take as input a formal context and an implication basis respectively. These algorithms run in polynomial time in the size of the input and the size of the output. When the input is an implication basis –actually the canonical basis in a concept lattice– we derive a condition such that a canonical basis corresponds to a distributive $\vee$-semilattice. However, this condition is necessary but not sufficient.

Then many directions are possible for future work. We have still to study how to deal with the TRIPLE condition when the input and output are an implication basis. Moreover, we should also work on the minimality of the output distributive lattices. The median graphs which are returned by the algorithms verify the properties (i) $L\backslash\bot$ can be order-embedded into a semilattice SL, and (ii) there is an isomorphism between $(\mathcal{J}(L), \leq)$ and $(\mathcal{J}(SL \cup \{\bot\}), \leq)$. The property (i) seems to be natural in applications. In phylogeny problems for example, one can be interested in adding latent vertices representing the species not yet observed, while the relative order of the observed species should not be changed. The second property can be considered as an extension of the first one, as all elements in a lattice can be inferred from irreducible elements. In this way, the invariance of the poset of $\vee$-irreducible elements ensures that the "core" of the data under study does not change.

Finally, we have also to work on the minimality of the output structures. We have shown and discussed the lack of a minimum solution when the two properties (*i*) and (*ii*) are satisfied. Thus the existence and the way how to reach such a minimum solution if any remains to be more deeply investigated.

References

[1] C. Carpineto, G. Romano, Concept Data Analysis: Theory and Applications, John Wiley & Sons, 2004.

[2] J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, G. Dedene, Formal Concept Analysis in Knowledge Processing: A Survey on Applications, Expert Systems with Applications 40 (16) (2013) 6538–6560.

[3] L. Vigilant, R. Pennington, H. Harpending, T. D. Kocher, A. C. Wilson, Mitochondrial DNA sequences in Single Hairs from a Southern African population, Proceedings of the National Academy of Sciences 86 (23) (1989) 9350–9354.

[4] H.-J. Bandelt, P. Forster, A. Röhl, Median-Joining Networks for Inferring Intraspecific Phylogenies, Molecular Biology and Evolution 16 (1) (1999) 37–48.

[5] H.-J. Bandelt, V. Macaulay, M. Richards, Median Networks: Speedy Construction and Greedy Reduction, One Simulation, and Two Case Studies from Human mtDNA, Molecular Phylogenetics and Evolution 16 (1) (2000) 8–28.

[6] M. Bernt, D. Merkle, M. Middendorf, Using Median Sets for Inferring Phylogenetic Trees, Bioinformatics 23 (2) (2007) e129–e135.

[7] H.-J. Bandelt, J. Hedlíková, Median algebras, Discrete mathematics 45 (1) (1983) 1–30.

[8] S. P. Avann, Median Algebras, Proceedings of the American Mathematical Society 12 (1961) 407–414.

[9] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, 1999.

[10] U. Priss, Concept Lattices and Median Networks, in: L. Szathmary, U. Priss (Eds.), Proceedings of The Ninth International Conference on Concept Lattices and Their Applications (CLA), Vol. 972 of CEUR Workshop Proceedings 972, 2012, pp. 351–354.

[11] U. Priss, Representing Median Networks with Concept Lattices, in: H. D. Pfeiffer, D. I. Ignatov, J. Poelmans, N. Gadiraju (Eds.), Proceedings of the 20th International Conference on Conceptual Structures (ICCS, Vol. 7735 of Lecture Notes in Computer Science 7735, Springer, 2013, pp. 311–321.

[12] A. Gély, M. Couceiro, L. Miclet, A. Napoli, Steps in the Representation of Concept Lattices and Median Graphs, in: F. J. Valverde-Albacete, M. Trnecka (Eds.), Proceedings of the Fifthteenth International Conference on Concept Lattices and Their Applications (CLA), CEUR Workshop Proceedings 2668, 2020, pp. 247–258.

[13] A. Gély, M. Couceiro, A. Napoli, Steps Towards Achieving Distributivity in Formal Concept Analysis, in: D. I. Ignatov, L. Nourine (Eds.), Proceedings of the Fourteenth International Conference on Concept Lattices and Their Applications (CLA), CEUR Workshop Proceedings 2123, 2018, pp. 105–116.

[14] G. Birkhoff, O. Frink, Representations of lattices by sets, Transactions of the American Mathematical Society 64 (2) (1948) 299–316.

[15] B. A. Davey, H. A. Priestley, Introduction to Lattices and Order, Cambridge university press, 2002.

[16] N. Caspard, B. Leclerc, B. Monjardet, Finite Ordered Sets: Concepts, Results and Uses, Cambridge University Press, 2012.

[17] B. Ganter, O. Sergei, Conceptual Exploration, Springer, 2016.

[18] J. L. Guigues, V. Duquenne, Famille minimale d'implications informatives résultant d'un tableau de données binaires, Mathématiques et Sciences Humaines 24 (95) (1986) 5–18.

[19] J. B. Nation, A. Pogel, The Lattice of Completions of an Ordered Set, Order 14 (1) (1997) 1–7.

[20] A. Gély, M. Couceiro, A. Napoli, Embedding Median Graphs into Minimal Distributive Join-Semi-Lattices, in: Proceedings of the International Workshop "New Frontiers in Mining Complex Patterns" (co-located with ECML-PKDD), 2019, http://www.di.uniba.it/ loglisci/NFMCP2019/program.html.

[21] R. P. Anstee, M. Farber, Characterizations of Totally Balanced Matrices, Journal of Algorithms 5 (2) (1984) 215–230.

[22] F. Brucker, P. Préa, Totally Balanced Formal Context Representation, in: J. Baixeries, C. Sacarea, M. Ojeda-Aciego (Eds.), Proceedings of the 13th International Conference on Formal Concept Analysis, Lecture Notes in Computer Science 9113, Springer, 2015, pp. 169–182.

[23] J. Demetrovics, L. Libkin, I. B. Muchnik, Functional Dependencies in Relational Databases: A Lattice Point of View, Discrete Applied Mathematics 40 (2) (1992) 155–185.

[24] B. Monjardet, The presence of lattice theory in discrete problems of mathematical social sciences. Why, Mathematical Social Sciences 46 (2) (2003) 103–144.