Correspondence-Based Lattice Similarity Measure

Florent Domenach and Zeynab Rajabi

Abstract This paper is in the formal concept analysis framework, an algebraic hierarchisation method of data based on the notion of extent/intent, i.e. of maximally shared attributes and objects. Here we present a correspondence-based similarity measure between two formal concept lattices, and compare it to results of a previous paper which introduced a structure-based dissimilarity measure. We define an expressive model using correspondences between objects and between attributes of the two lattices. A key point of our approach is that the correspondences may not be mappings and may associate each object (resp. attribute) of one lattice with several objects (resp. attributes) of another one.

1 Introduction

Lattices are polymorphic objects: as ordered sets, lattices are natural generalizations of many different ordered structures like trees, weak trees or pyramids. They are also the underlying structure in the Formal Concept Analysis framework articulating the duality between intent and extent. As such, they can be

Florent Domenach
Akita International University, Yuwa, Akita-city 010-1292 Japan,
✉ fdomenach@aiu.ac.jp

Zeynab Rajabi
✉ zr.1995@yahoo.com

ARCHIVES OF DATA SCIENCE (ONLINE FIRST) DOI 10.5445/KSP/1000058749/04
KIT SCIENTIFIC PUBLISHING ISSN 2363-9881
Vol. 2, No. 1, 2017

seen as an algebraic method of hierarchisation of data based on maximally shared attributes and objects.

The aim of this paper is to investigate a novel approach to evaluate the similarity between two lattices. Although there is an extensive literature on similarity between (directed) graphs (Ullmann, 1976; Goldsmith and Davenport, 1990; Raymond et al, 2002; Champin and Solnon, 2003), to the author's knowledge surprisingly little work exists for similarity of concept lattices (Domenach, 2015), particularly when considered as a specific case of digraphs. Using an approach similar in nature to Champin and Solnon (2003), the method introduced here is based on an expressive model using correspondences between objects and between attributes, correspondences which may not be one-to-one or onto. Each object (attribute) of the first lattice may be associated with any number of objects (attributes) of the second lattice.

This paper is organized as follows: In Sect. 2, the fundamental definitions and principles of Formal Concept Analysis are recalled. Section 3 states the problem central to this article together with previous attempts to solve it. Section 4 is the original work of this paper and describes our new proposed approach to tackle concept lattices similarity before illustrating it with several examples. Section 6 concludes the paper.

2 Formal Concept Analysis

2.1 Introduction

We recall here the standard Formal Concept Analysis (FCA) notations and we refer readers to Ganter and Wille (1999) for details and proofs on FCA and Caspard et al (2012) for results on lattices as ordered sets.

A *formal context* (G,M,I) is defined as a set G of objects, a set M of attributes, and a binary relation $I \subseteq G \times M$. $(g,m) \in I$ means that "the object g is related with the attribute m through the relation I". Table 1 shows an example of a formal context where the set of objects is all natural numbers less than or equal to 10, and the set of attributes are the properties of the set of numbers considered (*Composite*, *Even*, *Odd*, *Prime*, *Square*).

Table 1 Context of numbers with some basic properties

	Composite	Even	Odd	Prime	Square
1			X		X
2		X		X	
3			X	X	
4	X	X			X
5			X	X	
6	X	X			
7			X	X	
8	X	X			
9	X		X		X
10	X	X			

Two derivation operators can be defined on sets of objects and sets of attributes as follows, $\forall O \subseteq G, A \subseteq M$:

$$O' = \{m \in M : \forall g \in O, (g,m) \in I\}$$
$$A' = \{g \in G : \forall m \in A, (g,m) \in I\}$$

These two operators $(\cdot)'$ define a Galois connection between the power set of objects $\mathscr{P}(G)$ and the power set of attributes $\mathscr{P}(M)$ since both operators are antitone and satisfy, $\forall A \subseteq G, B \subseteq M : A \subseteq B' \iff B \subset A'$. A pair $(O,A), O \subseteq G, A \subseteq M$, is a *formal concept* iff $O' = A$ and $A' = O$. O is called the *extent* and A the *intent* of the concept. The composition of these two operators $(.)''$ forms a closure operator on $\mathscr{P}(G)$ (resp. $\mathscr{P}(M)$), and the Galois connection creates a (dual) isomorphism between the closed sets of $\mathscr{P}(G)$ and $\mathscr{P}(M)$.

Going back to our toy example of Table 1, consider the set of numbers $\{3,5\}$ - they are both *Odd* and *Prime*, so

$$\{3,5\}' = \{Odd, Prime\}$$

However, $\{7\}$ is also an odd prime number:

$$\{Odd, Prime\}' = \{3,5,7\}$$

$(\{3,5,7\},\{Odd,Prime\})$ is a formal concept with extent $\{3,5,7\}$ and intent $\{Odd,Prime\}$. Another manner to express the notion of concept is that of maximal rectangles: $\{3,5,7\}$ have all the attributes of $\{Odd,Prime\}$, and are the only numbers less than 10 to be both *Odd* and *Prime*. Another example of a concept is $(\{1,9\},\{Odd,Square\})$, as $\{1,9\}' = \{Odd,Square\}$ and $\{Odd,Square\}' = \{1,9\}$.

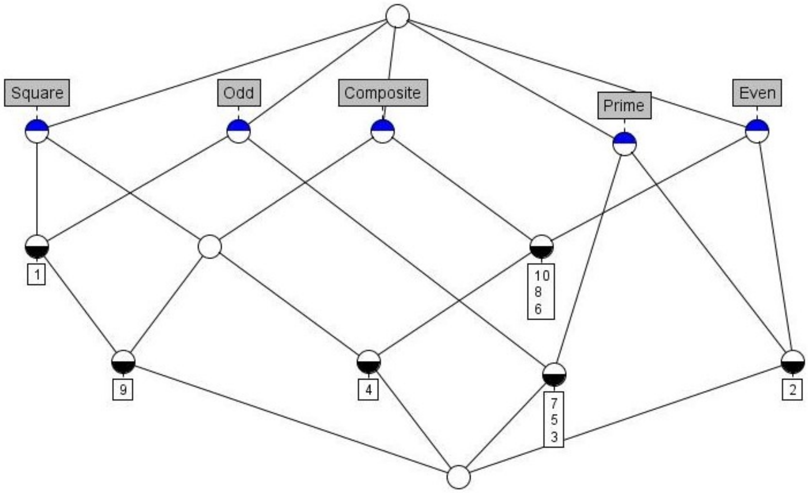

Fig. 1 Concept lattice associated with Table 1 with minimal labeling

2.2 Galois Lattice

The set of all formal concepts, ordered by inclusion of extents (or dually by inclusion of intents), *i.e.*, $(O_1,A_1) \leq (O_2,A_2)$ if $O_1 \subseteq O_2$ (or dually $A_2 \subseteq A_1$), forms a complete lattice (Barbut and Monjardet, 1970), called *concept lattice* denoted by $\mathbb{L} = \mathfrak{B}(G,M,I)$. A Hasse diagram can be associated with the concept lattice as the graph of the cover relation. Concept (O_1,A_1) is covered by concept (O_2,A_2), $(O_1,A_1) \prec (O_2,A_2)$, when there is no concept (O_3,A_3) such that $(O_1,A_1) < (O_3,A_3) < (O_2,A_2)$. In the Hasse diagram, each concept of the lattice is represented as a vertex in the plane and edges that go upward from (O_1,A_1) to (O_2,A_2) whenever $(O_1,A_1) \prec (O_2,A_2)$.

The concept lattice associated with our toy example of Table 1 is shown in Fig. 1, with minimal labelling: every vertex is a concept which inherits attributes that are above it and objects that are below it, and the edges represent the cover relation. Consider the vertex labeled with $\{1\}$: it represents the concept with extent $\{1,9\}$ (all the objects below the vertex) and intent $\{Square, Odd\}$ (all the attributes above the vertex). All the figures included in this paper were drawn using ConExp software[1] (Yevtushenko, 2000).

[1] Available at `http://conexp.sourceforge.net/`

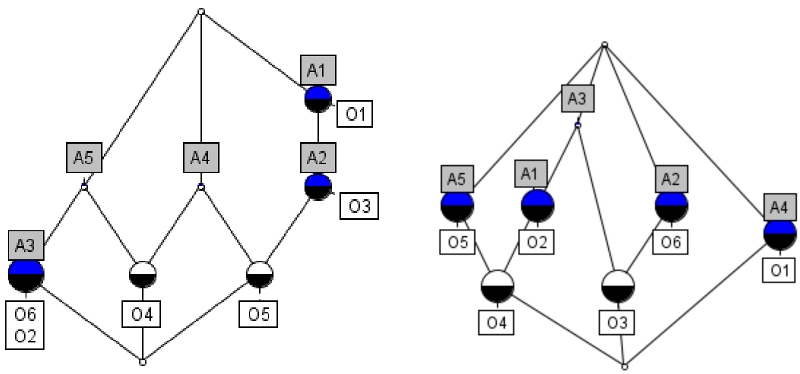

Fig. 2 Toy lattices $\mathbb{L}_1$ and $\mathbb{L}_2$

A well-known result on lattices is that the set of all lattices $\mathscr{L}$ is also a lattice, and the order associated with the lattice $\mathscr{L}$ is denoted by $\sqsubseteq$ (see for more detailed results Monjardet, 2003; Caspard and Monjardet, 2004).

3 Similarity Measures Between Lattices

3.1 Problem Statement

The problem considered in this paper is the quantification of the similarity between two given lattices, possibly defined on different sets of objects and attributes. Ideally such measure should take on large values for similar lattices and small values for very dissimilar lattices. Figure 2 shows two such toy lattices. The question then is how close are these two lattices? Since FCA is particularly useful in information retrieval and knowledge representation, studying similarity measures are particularly relevant for the comparison of lattices.

3.2 Existing Similarity Measures

We defined in Domenach (2015) a dissimilarity measure based on the structure of both lattices and normalized by the width of the lattices. Given two lattices

$\mathbb{L}_1$ and $\mathbb{L}_2$, $\mathbb{L}_1 = \mathfrak{B}(G,M,I_1)$ and $\mathbb{L}_2 = \mathfrak{B}(G,M,I_2)$, a dissimilarity measure on objects between $\mathbb{L}_1$ and $\mathbb{L}_2$ is defined as follows:

- Consider the $\frac{n(n-1)}{2}$ pairs of different objects $\{r,s\}, \forall r,s \in G$.
- We create the binary $\frac{n(n-1)}{2} \times n$ matrix $\mathbf{M}_{\mathbb{L}_1}$ with $m_{ik\mathbb{L}_1} = 1$ for objects $i = \{r,s\}$ if $\{r,s\}'' \neq \{r,s,k\}''$, i.e. if $\{r,s\}$ and $\{r,s,k\}$ have different closure in $\mathbb{L}_1$, with $r,s,k \in G$, and $m_{ik\mathbb{L}_1} = 0$ otherwise.
- Similarly, we create the matrix $\mathbf{M}_{\mathbb{L}_2}$ on pairs of objects of $\mathbb{L}_2$.

The dissimilarity measure is then defined as

$$d = \frac{||\mathbf{M}_{\mathbb{L}_1} - \mathbf{M}_{\mathbb{L}_2}||}{||\mathbf{M}_{\mathbb{L}_1}|| + ||\mathbf{M}_{\mathbb{L}_2}||}$$

where $||\mathbf{M}|| = \Sigma_i \Sigma_k |m_{ik}|$ is the L_1 norm of matrix $\mathbf{M}$. This measure is normalized, bounded in $[0,1]$, with $d = 0$ if the two lattices are identical and $d = 1$ if and only if no concept of $\mathbb{L}_1$ intersects a concept of $\mathbb{L}_2$, i.e. $\forall C_1 \in \mathbb{L}_1, C_2 \in \mathbb{L}_2, C_1 \cap C_2 = \emptyset$. Moreover, this dissimilarity does not depend on object labels since it refers to pairs of objects only.

Concept lattices create a hierarchy of the dual extent/intent of concepts, which is lost if objects and attributes are considered separately. Although a dual measure can be similarly defined on attributes, this dissimilarity measure does not take into account this fundamental aspect of concept lattices.

4 Proposed Similarity Measure

Jaccard (1901) created a simple similarity measure on sets defined as the ratio between the commonality of the elements over their union, which was later on generalized by Tversky (1977). When applied to concept lattices, the Jaccard measure can be written as:

$$sim(\mathbb{L}_1, \mathbb{L}_2) = \frac{f(descr(\mathbb{L}_1) \sqcap descr(\mathbb{L}_2))}{f(descr(\mathbb{L}_1) \sqcup descr(\mathbb{L}_2))} \; \forall \mathbb{L}_1, \mathbb{L}_2 \in \mathscr{L} \qquad (1)$$

The focus in this paper is on the case where f is the cardinality map, but the results can be easily extended to any function positive and monotonic non-decreasing with respect to the order on $\mathscr{L}$ ($\mathbb{L}_1 \sqsubseteq \mathbb{L}_2$ implies $f(\mathbb{L}_1) \leq f(\mathbb{L}_2)$). *descr* is a description function, which can be seen as a coding of a lattice allowing for a comparison between the two lattices. In the following sections,

Table 2 Example of an arbitrary correspondence c_o between objects of the toy lattices of Fig. 2, each row being an element of G_1 and each column an element of G_2

c_o	o_1	o_2	o_3	o_4	o_5	o_6
o_1				X		
o_2				X		
o_3						
o_4	X			X		
o_5			X	X		
o_6				X		

we first define the commonality of the description between the two lattices (Sect. 4.2), our numerator, before stating their union in Sect. 4.3, our denominator.

4.1 Correspondence

In order to define a similarity measure between two lattices $\mathbb{L}_1 = \mathfrak{B}(G_1, M_1, I_1)$ and $\mathbb{L}_2 = \mathfrak{B}(G_2, M_2, I_2)$, we first define an (arbitrary) correspondence c_o (resp. c_a) that matches objects (resp. attributes) in order to identify their common features. Those correspondences can be seen either as expert knowledge, matching features from a lattice to the other, or as an optimization problem, finding the best possible matching. Formally, $c_o \in G_1 \times G_2$ is a binary relation between the objects of $\mathbb{L}_1$ and the objects of $\mathbb{L}_2$. It is not a mapping between G_1 and G_2 as any object can have zero, one or more associated objects. Table 2 shows an example of a correspondence between objects of the two toy concept lattices of Fig. 2.

Given a correspondence c_o between G_1 and G_2 associating zero or more objects of $\mathbb{L}_2$ with each object of $\mathbb{L}_1$, we define the image of an object $o_1 \in G_1$ as:

$$c_o(o_1) = \{o_2 \in G_2 : (o_1, o_2) \in c_o\}$$

This definition can be extended to any set of objects $O_1 \subseteq G_1$ as the Cartesian product of the images of each element of O_1:

$$c_o(O_1) = \{\{y_1, y_2, ...\}, y_i \in c_o(o_i)\ \forall o_i \in O_1\} \text{ and } c_o(\emptyset) = \emptyset$$

For example, using the correspondence of Table 2, $c_0(\{o_4, o_5\}) = \{\{o_1, o_3\}, \{o_1, o_4\}, \{o_3, o_4\}, \{o_4\}\}$. Similar definitions of the image of an attribute or a set of attributes are used with the correspondence c_a that matches attributes between M_1 and M_2.

4.2 Defining Common Descriptions

In order to define the common description between two concept lattices, first we need to define how much of the information contained in $\mathbb{L}_1$ is represented in $\mathbb{L}_2$. We define this information as the ratio for each concept of $\mathbb{L}_1$ to be present, at least partially, in $\mathbb{L}_2$ through the correspondence c_o.

4.2.1 Description on Objects

Consider the concept $\lambda = (O_1, A_1) \in \mathbb{L}_1$. The description on objects of the concept λ from $\mathbb{L}_1$ to $\mathbb{L}_2$ depending on the correspondence c_o, denoted by $descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda)$, is the union of $(O_1, a_1), a_1 \in A_1$, such that a_1 is part of a concept of $\mathbb{L}_2$ that contains an image by c_o of O_1. Informally, it corresponds to the information contained in the concept λ based on its set of objects that is present, at least partially, in $\mathbb{L}_2$ through c_o. Formally,

$$descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda) = \{(O_1, a_1), a_1 \in A_1,\ \exists X_1 \in c_o(O_1) : a_1 \in X_1'\}$$

The overall description on objects from $\mathbb{L}_1$ to $\mathbb{L}_2$ is the union of the descriptions of each concept:

$$descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2} = \bigcup_{\lambda \in \mathbb{L}_1} descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda)$$

The common description between two concept lattices is then the union of the descriptions from one lattice to the other, i.e.

$$descr^{c_o}(\mathbb{L}_1) \sqcap descr^{c_o}(\mathbb{L}_2) = descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2} \cup descr^{c_o}_{\mathbb{L}_2 \to \mathbb{L}_1}$$

Continuing with our toy lattices example of Fig. 2, and using the correspondence c_o of Table 2, we have $descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2}(\{o_3, o_5\}, \{a_1, a_2\}) = \{(\{o_3, o_5\}, \{a_1\}), (\{o_3, o_5\}, \{a_2\})\}$ as $c_o(\{o_3, o_5\}) = \{\{o_3\}, \{o_4\}\}$, $a_1 \in \{o_4\}'$ and $a_2 \in \{o_3\}'$. However, $descr^{c_o}_{\mathbb{L}_1 \to \mathbb{L}_2}(\{o_4, o_5\}, \{a_4\}) = \emptyset$ since $c_o(\{o_4, o_5\}) = \{\{o_1, o_3\}, \{o_1, o_4\}, \{o_3, o_4\}, \{o_4\}\}$ but the only concept of $\mathbb{L}_2$ containing a_4 is $(\{o_1\}, \{a_4\})$.

4.2.2 Description on Attributes

Given a correspondence c_a associating zero or more attributes of M_2 with each attribute of M_1, we can dually define the common (attribute) description from $\mathbb{L}_1$ onto $\mathbb{L}_2$ for a concept $\lambda = (O_1, A_1) \in \mathbb{L}_1$ as:

$$descr^{c_a}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda) = \{(o_1, A_1), o_1 \in O_1, \exists Y_1 \in c_a(A_1) : o_1 \in Y_1'\}$$

and the attribute description from $\mathbb{L}_1$ to $\mathbb{L}_2$ as: $descr^{c_a}_{\mathbb{L}_1 \to \mathbb{L}_2} = \bigcup_{\lambda \in \mathbb{L}_1} descr^{c_a}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda)$

4.2.3 Description on Objects and Attributes

None of the two previous definitions of descriptions on objects and on attributes are entirely satisfying as they consider objects and attributes separately. In order to take into account the dual nature of concept lattices, these definitions of descriptions, either on objects or on attributes, lead to a unified description on both dimensions. The description from $\mathbb{L}_1$ to $\mathbb{L}_2$ on objects and attributes is a combination of the description on objects and the description on attributes, i.e. $\forall \lambda = (O_1, A_1) \in \mathbb{L}_1$:

$$\begin{aligned} descr^{c_o,c_a}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda) = \{&(O_1, a_1),\, a_1 \in A_1,\, \exists X_1 \in c_o(O_1),\, \exists y_1 \in c_a(a_1) :\, y_1 \in X_1'\} \\ &\cup \{(o_1, A_1),\, o_1 \in O_1,\, \exists Y_1 \in c_a(A_1),\, \exists x_1 \in c_o(o_1),\, z_1 \in Y_1'\} \end{aligned} \tag{2}$$

And, similarly, the description on objects and attributes from $\mathbb{L}_1$ to $\mathbb{L}_2$ is defined as:

$$descr^{c_o,c_a}_{\mathbb{L}_1 \to \mathbb{L}_2} = \bigcup_{\lambda \in \mathbb{L}_1} descr^{c_o,c_a}_{\mathbb{L}_1 \to \mathbb{L}_2}(\lambda)$$

Continuing with our toy example of Fig. 2, and using c_o of Table 2 and c_a as the identity correspondence ($\forall i, c_a(a_i) = a_i', a_i \in M_1, a_i' \in M_2$), we have

$$\begin{aligned} descr^{c_o,c_a}_{\mathbb{L}_1 \to \mathbb{L}_2}(\{o_3, o_4\}, \{a_1, a_5\}) = \\ \{(\{o_3, o_4\}, \{a_1\}), (\{o_3, o_4\}, \{a_2\}), (\{o_4\}, \{a_1, a_2\})\}. \end{aligned}$$

The common description of $\mathbb{L}_1$ and $\mathbb{L}_2$, used as the numerator in equation 1, is the union of the description from $\mathbb{L}_1$ to $\mathbb{L}_2$ and the description from $\mathbb{L}_2$ to $\mathbb{L}_1$. This set contains all the features from both $\mathbb{L}_1$ and $\mathbb{L}_2$ that are partially matched by the correspondences c_o and c_a.

Definition 1. The common description of two lattices $\mathbb{L}_1$ and $\mathbb{L}_2$, denoted by $descr(\mathbb{L}_1) \sqcap descr(\mathbb{L}_2)$, is defined as:

$$descr(\mathbb{L}_1) \sqcap descr(\mathbb{L}_2) = descr^{c_o,c_a}_{\mathbb{L}_1 \to \mathbb{L}_2} \cup descr^{c_o,c_a}_{\mathbb{L}_2 \to \mathbb{L}_1} \tag{3}$$

4.3 Union of Descriptions

In order to complete our definition of the Jaccard similarity between two lattices, used as the denominator in equation 1, we need to define the union of the descriptions of those lattices.

Definition 2. The union of descriptions of two lattices $\mathbb{L}_1$ and $\mathbb{L}_2$, denoted as $descr(\mathbb{L}_1) \sqcup descr(\mathbb{L}_2)$, is defined as:

$$\begin{aligned} descr(\mathbb{L}_1) \sqcup descr(\mathbb{L}_2) = & \bigcup_{(O_1,A_1) \in \mathbb{L}_1} \{ \bigcup_{a_1 \in A_1} \{(O_1, a_1)\} \cup \bigcup_{(o_1 \in O_1)} \{(o_1, A_1)\} \} \\ & \cup \bigcup_{(O_2,A_2) \in \mathbb{L}_2} \{ \bigcup_{a_2 \in A_2} \{(O_2, a_2)\} \cup \bigcup_{(o_2 \in O_2)} \{(o_2, A_2)\} \} \end{aligned} \tag{4}$$

4.4 Splits

An issue of this correspondence based approach is that c_o and c_a are binary relations, not mappings. So any object (attribute) of $\mathbb{L}_1$ or $\mathbb{L}_2$ can have more than one image. Consider the extreme case where every object / attribute of $\mathbb{L}_1$ is related to every other object / attribute of $\mathbb{L}_2$. Although uninformative, the similarity, as defined in equation 1, will be artificially high.

Splits are defined when an object or attribute has more than one image in c_o or c_a. Informally, splits quantify the lack of precision in the correspondences c_o and c_a. The more images an object (resp. an attribute) has through c_o (resp. c_a), the less informative it is.

We can now revisit our similarity measure of equation 1 by taking splits into account as we want to have a similarity measure that will be decreasing as the number of splits is increasing:

Definition 3. The correspondence-based lattice similarity measure, depending on the correspondences c_o and c_a between objects and attributes respectively, is defined as:

$$sim_{c_o,c_a}(\mathbb{L}_1,\mathbb{L}_2) = \frac{f(descr(\mathbb{L}_1)\sqcap descr(\mathbb{L}_2)) - g(splits(c_o)\cup splits(c_a))}{f(descr(\mathbb{L}_1)\sqcup descr(\mathbb{L}_2))} \quad (5)$$

with f and g two positive and non-decreasing mappings (here cardinality mappings).

The similarity $sim_{c_o,c_a}(\mathbb{L}_1,\mathbb{L}_2)$ is a similarity measure as we have, trivially, $sim_{c_o,c_a}(\mathbb{L}_1,\mathbb{L}_1) = 1$ with diagonal c_a, c_o. However, it is not normalized because of possible splits, and so can become negative.

Although different definitions of the splits can be used for this similarity measure, we focused on two specific cases where:

- $split_1$ is the number of rows and columns of c_o and c_a with more than one image, i.e. $split_1(c_o) = \{(o,c_o(o)) : o \in G_1\cup G_2, |c_o(o)| > 1\}$ and $split_1(c_a) = \{(a,c_a(a)) : a \in M_1\cup M_2, |c_a(a)| > 1\}$.
- $split_2$ is the number of images minus one in c_o and c_a of each object and attributes, i,e, $\Sigma_{x\in G_1\cup G_2}(|c_o(x)|-1) + \Sigma_{y\in M_1\cup M_2}(|c_a(y)|-1)$.

While being simpler, the mapping $split_1$ tends to be too conservative and does not sufficiently penalize correspondences with multiple images, as we will illustrate through a simple example in the next section.

5 Examples

In this section we present some examples of concept lattices together with the correspondences on objects and on attributes to illustrate our similarity measure.

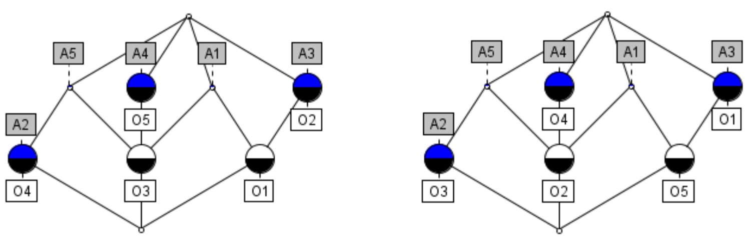

Fig. 3 Concepts lattices $\mathbb{L}_1$ (left) and $\mathbb{L}_2$ (right) obtained by permutation

Table 3 Correspondences c_o (representing a permutation on objects, left) and c_a (right) for the lattices of Fig. 3

c_o	o_1	o_2	o_3	o_4	o_5
o_1		X			
o_2			X		
o_3				X	
o_4					X
o_5	X				

c_a	a_1	a_2	a_3	a_4	a_5
a_1	X				
a_2		X			
a_3			X		
a_4				X	
a_5					X

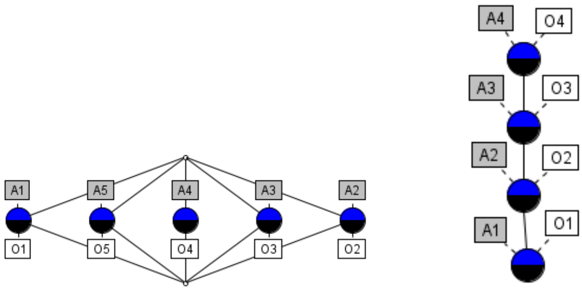

Fig. 4 Antichain lattice $\mathbf{M}_5$ (left) and linear order lattice $\mathbf{C}_4$ (right)

Table 4 Correspondences c_o (left) and c_a (right) for the lattices of Fig. 4

c_o	o_1	o_2	o_3	o_4
o_1	X			
o_2		X		
o_3			X	
o_4				X
o_5				

c_a	a_1	a_2	a_3	a_4
a_1	X			
a_2		X		
a_3			X	
a_4				X
a_5				

The examples were chosen in order to exemplify different aspects, strengths and weaknesses, of this new measure.

Our first example is a permutation example. Consider the lattices of Fig. 3: the lattice $\mathbb{L}_2$ was obtained by permuting the objects of $\mathbb{L}_1$. As one would expect, the two lattices are identical except for their object labeling - using the correspondences of Table 3, we indeed find that the similarity on objects and Attribute is 100%.

The second example is another extreme example, where we wanted to evaluate the behavior of the similarity measure when the lattices are strongly dissim-

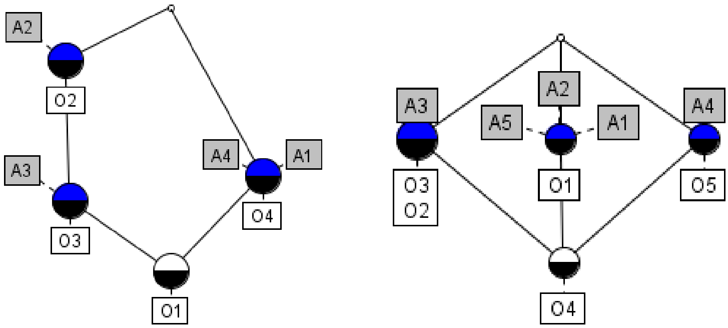

Fig. 5 Lattice $\mathbf{N}_5$ (left) and antichain lattice $\mathbf{M}_3$ (right)

Table 5 Correspondences c_o (left) and c_a (right) for the lattices of Fig. 5

c_o	o_1	o_2	o_3	o_4	o_5
o_1	X	X	X	X	X
o_2	X			X	
o_3	X	X	X	X	
o_4	X			X	X

c_a	a_1	a_2	a_3	a_4	a_5
a_1	X	X	X	X	X
a_2	X	X	X		X
a_3	X	X	X		X
a_4	X	X	X	X	X

ilar. The lattices of Fig. 4 are, on the left, the antichain lattice with 5 elements, where all the non-extremal concepts are not comparable to each other, and, on the right, the linear order where any two concepts are comparable. Using the correspondences of Table 4, we obtain a similarity of 33.3%.

The last example illustrates the impact of possible choices for the split function in equation 5. It is also another aberrant case, where each object and attribute in the correspondences c_o and c_a have multiple images (Table 5), while the lattices are quite dissimilar (Fig. 5). With function $split_1$, which counts the splits as the number of rows and columns having more than one image, the similarity is 48%. But when applying function $split_2$, which is the number of images minus one in c_o and c_a, the similarity become -31%. This example shows that the mapping $split_2$ may be more suitable compared to a more conservative $split_1$, which can be explained by the use of the existence operator in equation 2.

6 Conclusion and Perspectives

In this paper we introduced an adaptation of the Champin and Solnon (2003) similarity measure from directed graphs to the FCA framework. Based on the Jaccard measure, our similarity uses correspondences between objects and between attributes of the two concept lattices. Our similarity measure is able to capture the inter-related role of intent and extent, i.e. of objects and attributes between the two concept lattices considered.

Since our similarity measure relies on finding the best correspondences c_o and c_a, a key issue relies on providing efficient algorithms for their implementation. As f and g are non-decreasing, it is difficult to evaluate the change of similarity when increasing one of the correspondences. Another future investigation is the statistical evaluation of the similarity measure depending on the mappings f and g. We are also planning to analyze its behavior and correlate it in relation with existing similarity measures.

References

Barbut M, Monjardet B (1970) Ordres et classification: Algèbre et combinatoire (tome II). Hachette, Paris

Caspard N, Monjardet B (2004) Some lattices of closure systems on a finite set. Discrete Mathematics and Theoretical Computer Sciences 6:163–190, DOI 10.1016/S0166-218X(02)00209-3

Caspard N, Leclerc B, Monjardet B (2012) Finite ordered sets: Concepts, results and uses. 144, Cambridge University Press, Cambridge

Champin PA, Solnon C (2003) Measuring the similarity of labeled graphs. In: Ashley K, Bridge D (eds) Case-Based Reasoning Research and Development, vol 2689, Springer, Berlin, pp 80–95, DOI 10.1007/3-540-45006-8

Domenach F (2015) Similarity measures of concept lattices. In: Lausen B, Krolak-Schwerdt S, Böhmer M (eds) Data Science, Learning by Latent Structures, and Knowledge Discovery, Springer, Berlin, pp 89–99, DOI 10.1007/978-3-662-44983-7_8, URL `http://dx.doi.org/10.1007/978-3-662-44983-7_8`

Ganter B, Wille R (1999) Formal Concept Analysis: Mathematical Foundations. Springer, Berlin, DOI 10.1007/978-3-642-59830-2

Goldsmith TE, Davenport DM (1990) Assessing structural similarity of graphs. In: Schvaneveldt RW (ed) Pathfinder Associative Networks, Ablex Publishing Corp., Norwood, NJ, USA, pp 75–87

Jaccard P (1901) Étude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37:547–579, DOI 10.5169/seals-266450

Monjardet B (2003) The presence of lattice theory in discrete problems of mathematical social sciences. Why. Mathematical Social Sciences 46(2):103–144, DOI 10.1016/S0165-4896(03)00072-6

Raymond JW, Gardiner EJ, Willett P (2002) Heuristics for similarity searching of chemical graphs using a maximum common edge subgraph algorithm. Journal of Chemical Information and Computer Sciences 42(2):305–316, DOI 10.1021/ci010381f

Tversky A (1977) Features of similarity. Psychological Reviews 84(4):327–352, DOI 10.1037/0033-295X.84.4.327

Ullmann JR (1976) An algorithm for subgraph isomorphism. Journal of the ACM 23(1):31–42, DOI 10.1145/321921.321925

Yevtushenko SA (2000) System of data analysis "concept explorer". In: Proceedings of the 7th National Conference on Artificial Intelligence KII-2000, pp 127–134