

---

# A lattice structure in statistical implicative analysis

Authors : Fidy Andrianarivony, Elinambinina Rajaonarifara, Stefana Tabera Tsilefa, Angelo Raherinirina, Jean-Claude Régnier

May 11, 2024



# Outline

- 1 Background
- 2 Lattice structure
- 3 Statistical implication analysis
- 4 Constructing lattice from implicative graph
- 5 Illustration
- 6 Future work

---

## Background

# Background

- The concept of lattice has begun on 1949 (Birkhoff, 1949)
- It has been used in different fields: biology, computer sciences, other branch of mathematics.
- No significant application of lattice in statistical logics
- Statistical implication analysis: approach used in data mining to find relationship of implication between different variables.
- In this study, constructing a lattice structure from implicative graph

---

## Lattice structure

# Lattice structure

## A finite lattice

An ordered set  $(E, \leq)$  is a finite lattice if for each pair of elements of  $E$ , we have an infimum and a supremum. In particular, the infimum of  $E$ , if it exists, is denoted by  $\bigwedge_{x \in E} E = \perp$ . It is the unique minimal element of  $E$ . Similarly, the supremum of  $E$ , if it exists, is denoted by  $\bigvee_{x \in E} E = \top$ . It is the unique maximal element of  $E$ .

## A semilattice

In the case where  $(E, \leq)$  admits only the  $\perp$  infimum of  $E$ , then it is said to be a *lower-semilattice*. In the case where  $(E, \leq)$  admits only the supremum  $\top$  of  $E$ , then it is said to be a *upper-semilattice*.

---

## Statistical implication analysis

# Statistical implication analysis(SIA)

- The problem of SIA is to establish a rule in implicative form between two variables
- Notation for "*if x then y*":  $x \longrightarrow y$
- A theoretical framework is generated to model the existence of the relation ship
- It is based on measures that take into account the occurences of both variables



- 
- The widely used measure is the implication intensity( Gras and Regnier,2009)

$$\varphi(x \rightarrow y) = \begin{cases} 1 - \mathbb{P} \left[ \text{Card}(X \cap \bar{Y}) \leq n_{x \wedge \bar{y}} \right] & \text{if } n \neq n_y \\ 0 & \text{if } n = n_y \end{cases}$$

- It can be modeled from the Poisson distribution with parameter:  $\lambda = \frac{n_x \cdot n_{\bar{y}}}{n}$  and given by the formula:

$$\varphi(x \rightarrow y) = 1 - \sum_{k=0}^{n_{x \wedge \bar{y}}} \frac{\lambda^k}{k!} e^{-\lambda}$$

- A threshold  $\alpha$  is chosen in the interval  $]0.5, 1[$  to set the existence of the relationship when  $\varphi(x \rightarrow y) \geq \alpha$ .
- The retained rule from that measure is called: quasi-implication.

# Implicative graph

- An implicative graph can be generated
- It gives clearer visualisation of the relationship between variables
- If we have a set of binary variables  $E$ , it can be associated with the quasi-implication relation denoted by  $\rightarrow$  and defined by for  $x$  and  $y$  in  $E$ ,  $x \rightarrow y$  if and only if  $\varphi(x \rightarrow y) \geq \alpha$ .
- Example  
The set of variables is  $E = a, b, c, d, e, f, g$  and the binary relationship is as follows:  $\rightarrow = \{(e, c), (e, a), (e, f), (e, b), (c, a), (c, f), (b, a), (b, f), (g, d), (g, f), (a, f)\}$

---

The implicative graph corresponding to this is:

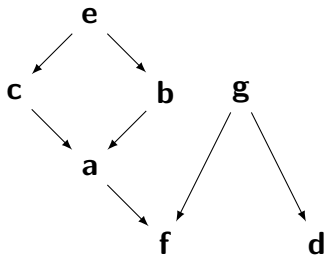


Figure: Implicative graph

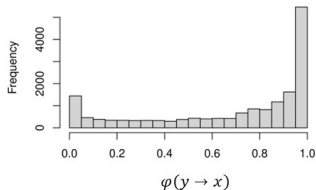
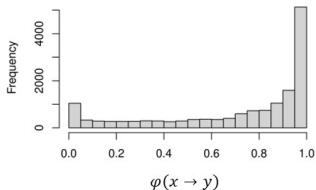
- 
- This relation is reflexive but it's not antisymmetric nor transitive.
  - So,  $E$ , associated with the quasi-implication is not an ordered set and obviously not a lattice. However, we need a transitive relation, so we admit the following proposition from (Gras and Regnier, 2009).

**Proposition 1** *For any  $\alpha \in ]0.5; 1[$ , by convention, if  $x \rightarrow y$  and  $y \rightarrow z$ , we have a transitive closure  $x \rightarrow z$  if, and only if  $\varphi(x \rightarrow z) \geq 0.5$ .*

- For our modeling purposes, we adopt the second proposal below:

**Proposition 2** *For two pairs of binary variables  $(x, y)$  and  $(z, t)$  of  $E^2$ , if  $x \rightarrow z$ ,  $x \rightarrow t$  and  $y \rightarrow z$ ,  $y \rightarrow t$  then  $x \rightarrow y$  or  $y \rightarrow x$ .*

- The proof of this proposition is proceeded by simulation and the result is shown in the figure below:



---

## Constructing lattice from implicative graph

Let  $E$  be a pre-order(reflexive and transitive) set,  $x$  and  $y$  two elements of  $E$  such that  $x \rightarrow y$ .

- Premise:  $x$  in the relation  $x \rightarrow y$  is called premise
- Consequent:  $y$  is called consequent
- Set of premises of an element  $x$ :  $P_x = \{t \in E, t \rightarrow x\}$
- Set of consequences of an element:  $G_x = \{t \in E, x \rightarrow t\}$



We have the following properties:

- $\forall x, y \in E P_x \subset P_y \Leftrightarrow x \longrightarrow y$
- For any  $x \in E$ , if there is no element equivalent to  $x$ , then  $P_x \cap G_x = \{x\}$ .
- For any  $x, y \in E$ , if  $P_y = G_x$  then the set  $P_y = G_x$  provided with the quasi-implication relation is a preordered set which admits as minimum  $x$  and as maximum  $y$ .

## Generating a lattice structure: lower semi-lattice

- Consider  $E$ , a set of binary variables associated with the relation quasi-implication. We define  $\tau_p$  as the set of all premises in  $E$  union the empty set. We have

$$\tau_p = \{P_x, x \in E\} \cup \{\emptyset\}$$

.  $\tau_p$  associated with the relation inclusion denoted by  $(\tau_p, \subset)$  is an ordered set and it defines a lower semi-lattice.

- **Proof:** Let  $a, b$  be two fixed elements of  $E$ . Consider the two sets  $A = \{t \in E, t \rightarrow a\}$  and  $B = \{t \in E, t \rightarrow b\}$  which are elements of  $\tau_p$ . There are three possible cases:
    - if  $A \subset B$  then  $A$  is the infimum of  $A$  and  $B$ ;
    - if  $A$  and  $B$  are not comparable and  $A \cap B = \emptyset$  then  $\emptyset$  is the infimum of  $A$  and  $B$ ;
    - if  $A$  and  $B$  are not comparable and  $A \cap B \neq \emptyset$  then there exists  $x \in A \cap B$  such that  $x \rightarrow a$  and  $x \rightarrow b$ . Now, according to Proposition 2, for all  $x \in A \cap B$ , there exists  $\vee x \in A \cap B$  such that  $x \rightarrow \vee x$ . Thus,  $A \cap B = \{t \in E, t \rightarrow \vee x\} \in \tau_p$  is the infimum of  $A$  and  $B$ . We have used the notation  $\vee x$  to indicate that it is the maximum element of  $x$ .
- So, in all cases, there exists an infimum for  $A$  and  $B$ .

- In the same way, we define  $\tau_c$  as the set of all consequences union  $E$

$$\tau_c = \{G_x, x \in E\} \cup \{E\}$$

- $(\tau_c, \subset)$  is an ordered set and defines an upper semi-lattice.

# Upper semi-lattice

- Proof: Let  $A, B$  be two elements of  $\tau_c$ . Let's show that there is an infimum and a supremum. We can imagine the following cases:
  - ① if  $A \subset B$  then  $B$  is supremum;
  - ② otherwise, we always have  $A \subset E$  and  $B \subset E$ , i.e.  $E$  is a majorant of  $A$  and  $B$ :
    - ① if there are no other majorants than  $E$ , then  $E$  is the supremum of  $A$  and  $B$ ;
    - ② if there exists a unique  $C \in \tau_c$  such that  $C \neq E$ , and  $A \subset C$  and  $B \subset C$ , then  $C$  is the supremum of  $A$  and  $B$ ;
    - ③ if there exists a family  $(C_i)_{i \in I}$  of  $\tau_c$  such that for any  $i \in I$ ,  $C_i$  contains both  $A$  and  $B$ , then we can show that  $\bigcap_{i \in I} C_i \in \tau_c$  is the supremum of  $A$  and  $B$ . Indeed, for any  $i \in I$ , we have  $A \subset C_i$  and  $B \subset C_i$ , hence  $A \subset \bigcap_{i \in I} C_i$  and  $B \subset \bigcap_{i \in I} C_i$ . The intersection of  $C_i$  is the smallest of the majorants of  $A$  and  $B$  because for any  $i \in I$ , we have  $\bigcap_{i \in I} C_i \subset C_i$ .

- We define  $\tau$  as the union of  $\tau_p$  and  $\tau_c$ :

$$\tau = \tau_p \cup \tau_c$$

- $\tau$ , associated with the relation  $\subset$  has the structure of lattice. We call it the lattice corresponding to the implicative graph

---

Illustration

# Illustration

- To illustrate the construction, we consider a set  $E = a, b, c, d, e, f, g$  with the corresponding implicative graph presented as follows:

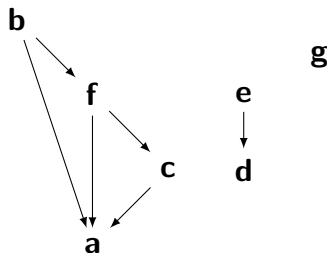


Figure: Implicative graph at threshold 0.6, generated by CHIC software



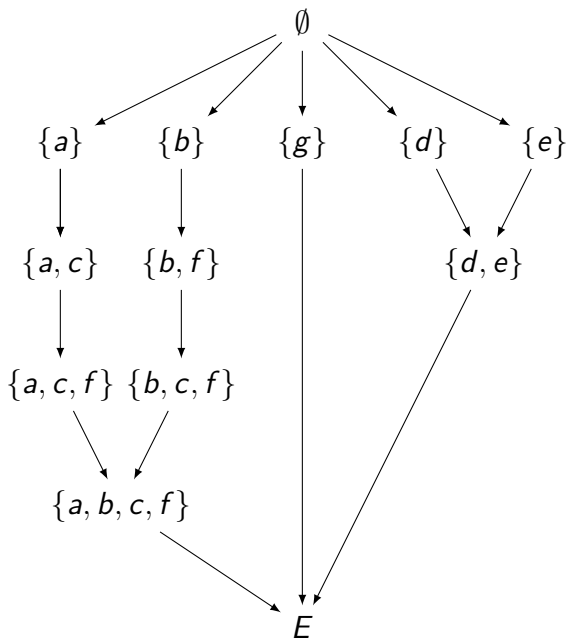
# Illustration

- The set of premises and consequents are:

Table: Sets of premises and consequents of  $E$ .

<b>Variables</b>	<b>Set of premisses</b>	<b>Set of consequents</b>
a	$\{a, b, c, f\}$	$\{a\}$
b	$\{b\}$	$\{a, b, c, f\}$
c	$\{b, c, f\}$	$\{a, c\}$
d	$\{d, e\}$	$\{d\}$
e	$\{e\}$	$\{d, e\}$
f	$\{b, f\}$	$\{a, c, f\}$
g	$\{g\}$	$\{g\}$

- The corresponding lattice structure is given by:



- This study established a link between two fields of mathematics and statistics, namely lattice theory and SIA.
- This is to optimize the information extracted by SIA.
- A mathematical method supported by a computational method to construct a lattice from the implicative graph is introduced.

---

Future work

## Limitations and future works

- We have not taken into account the minority case of implicative graphs that do not respect proposition 2
- A possible new approach is to develop an alternative method for constructing a lattice from all existing implicative graphs, without this restriction.
- This study did not focus on a concrete application of the resulting model.
- Application of this model in various fields of information extraction can be explored.

# Aknowledgement

- University of Fianarantsoa, Madagascar
- The LABEX ASLAN (ANR-10-LABX-0081), University of Lyon for the funding in the program: "Investissement d'Avenir" led by the "Agence Nationale de la Recherche (ANR)
- The "Ecole Normale Supérieure"
- The CREM laboratory