

# DU QUOTIDIEN AU PHILOSOPHIQUE

L'implication statistique illustre l'adage :  
**L'EXCEPTION CONFIRME LA REGLE**

La démarche statistique inductive cherche à transcender  
**LA CONTINGENCE EN LA NECESSITE**

Et l'objectif visé veut satisfaire :  
**LA SIGNIFICATION DU « TOUT » EST SUPERIEURE  
A LA SOMME DES SIGNIFICATIONS DE SES « PARTIES »**

D'où 4 concepts-clés en A.S.I. :  
**REGLE, INDUCTION, STRUCTURE, QUALITE INFORMATIVE**

# PANORAMA DU DEVELOPPEMENT DE L'ANALYSE STATISTIQUE IMPLICATIVE A PARTIR DE SITUATIONS FONDATRICES

Régis Gras, L.I.N.A.

3ème Rencontre Internationale sur l'Analyse Statistique Implicative  
Palerme 6-8 Octobre 2005



# **PLAN**

## **Introduction**

**§ 1 L'intensité d'implication classique (cas binaire)**

**§ 2 L'implication-inclusion**

**§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles**

**§ 4 Graphe d'implication**

**§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs**

**§ 6 Typicalité et Contribution des sujets et des variables supplémentaires**

**Conclusion**

# THEORIE DE L'IMPLICATION STATISTIQUE

DANS UN ENSEMBLE EXPERIMENTAL  
(ESPACE D'APPRENTISSAGE) on dispose :

- d'un ensemble  $V$  d'attributs ou caractères (variables)
- d'un ensemble  $E$  de sujets possédant ou non ces attributs

IL EXISTE AINSI SUR LES VARIABLES  
ET DANS  $E$  DES RELATIONS DU TYPE :

SI variable  $a$  = vraie (ou presque vraie)

ALORS

variable  $b$  = vraie (ou presque vraie) (REGLE :  $a \Rightarrow b$ )

# THEORIE DE L'IMPLICATION STATISTIQUE (suite)

....et aussi

**SI** règle **R = vraie** (ou presque) **ALORS**  
règle **S = vraie** (ou presque)

**IL EXISTE ALORS DES CONCEPTS**  
**QUI SONT ASSOCIES**

-----

Nous chercherons à l'aide de

**C.H.I.C.**

à mettre en évidence ces relations implicatives

L'expert du domaine se chargera

de les interpréter en terme de **concepts**

# PROBLEMES POSES PAR LE CROISEMENT $E \times V$

I

**Dégager** des règles entre les variables du type  $a \Rightarrow b$

II

**Mesurer** la qualité de ces règles (quasi-implications)

III

**Représenter** les structures de  $V$ , de  $V \times V$ , ...,  
issues des règles (graphe) et méta-règles (hiérarchie)

IV

**Mesurer** la qualité des structures obtenues

V

**Etudier** la responsabilité des éléments de  $E$   
ou des variables supplémentaires (sous-ensembles de  $E$ )  
par rapport aux structures

(cf. Gras, Kuntz, Briand, MSH 2001)

# LES ACTEURS DE LA MODELISATION

**E** : ensemble d'individus ou d'observations  $\text{card } E = n$

**V** : ensemble de variables ou de caractères ou d'attributs

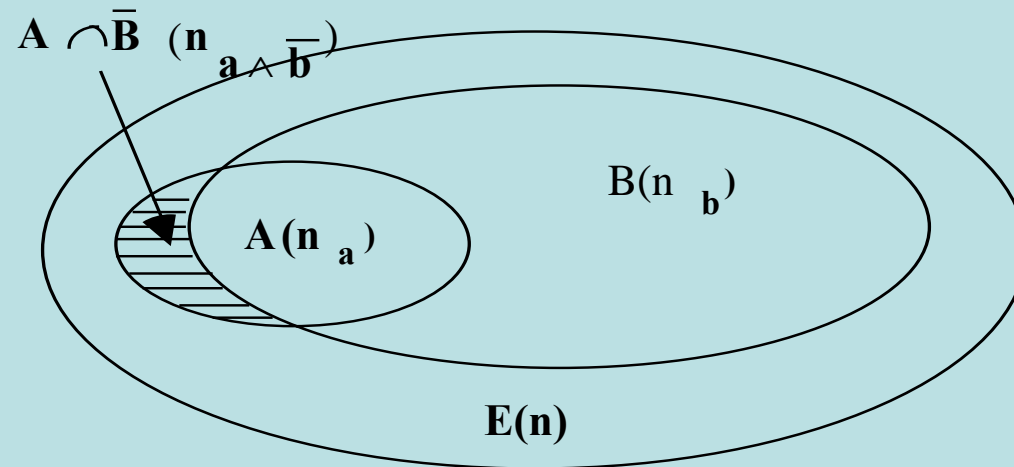
**A** : sous-ensemble de E des observations, exemples de réalisations de la variable a

**B** : sous-ensemble de E des observations, exemples de réalisations de la variable b

$$\text{card } A = n_a \quad \text{card } B = n_b$$

$A \cap \bar{B}$  sous-ensemble de E des observations, contre-exemples de l'implication  $a \Rightarrow b$

$$\text{Card } A \cap \bar{B} = n_{a \wedge \bar{b}}$$



# PLAN

Introduction

**§ 1 L'intensité d'implication classique (cas binaire)**

§ 2 L'implication-inclusion

§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles

§ 4 Graphe d'implication

§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs

§ 6 Typicalité et Contribution des sujets et des variables supplémentaires

Conclusion

# APPROCHE INTUITIVE

X et Y de même cardinaux que A et B

Si  $\text{card}(A \cap \bar{B})$  est invraisemblablement petit  
par rapport à la variable aléatoire  $\text{card}(X \cap \bar{Y})$ ,

on acceptera la QUASI-IMPLICATION  $a \Rightarrow b$

à la mesure de l'écart observé

**INTENSITE D'IMPLICATION = ETONNEMENT STATISTIQUE**

Comment le mesurer ?

*« Si la fréquence des coïncidences n'excède pas de façon significative la probabilité qu'on peut leur calculer en les attribuant au seul hasard à l'exclusion de relations causales cachées, nous n'avons certes aucune raison de supposer l'existence de telles relations. »*

H.Atlan, « A tort et à raison. Intercritique de la science et du mythe », Seuil, 1986

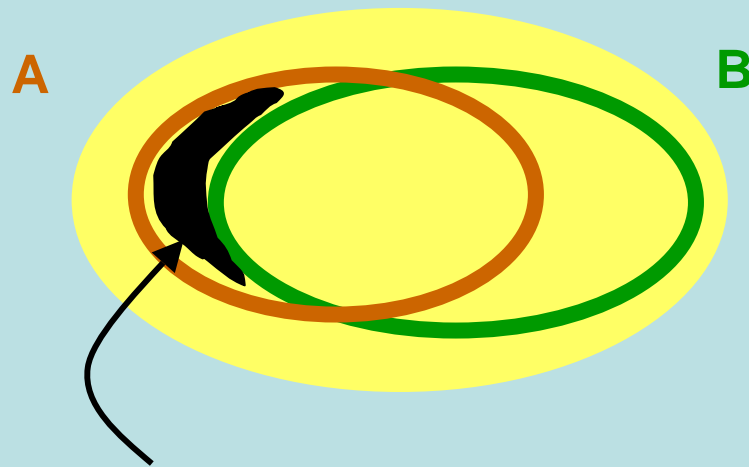
# INTENSITE D'IMPLICATION

$$a \Rightarrow b$$

**Modèle de tirage** : choix indépendant pour X et Y avec card X=card A et card Y=card B

Observations

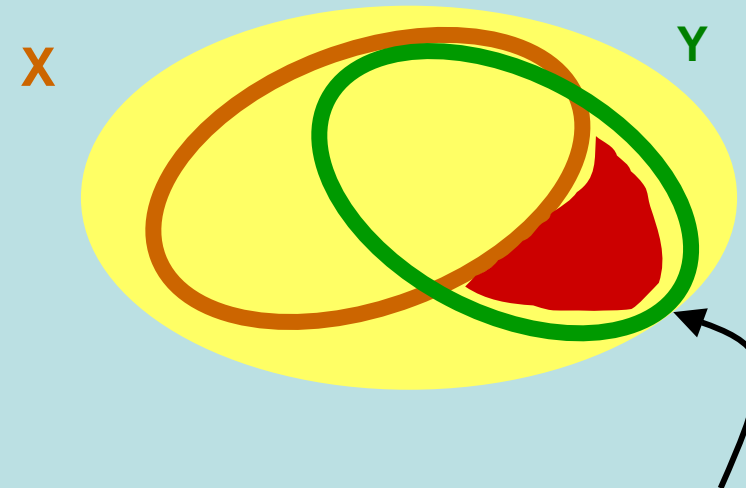
E



contre-exemples  
 $n(a \wedge \neg b)$

Modèle aléatoire

E



variable aléatoire  
 $N(a \wedge \neg b)$

La règle est **admissible** si  
 $Pr(N(a \wedge \neg b) < n(a \wedge \neg b))$   
est **petite**

# CHOIX DE MODELES PROBABILISTES POUR $\text{card}(X \cap \bar{Y})$

Plusieurs sont proposées en fonction  
du processus de tirage des éléments de E :

1° modèle hypergéométrique

2° modèle binomial de paramètres  $n, \frac{n_a n_{\bar{b}}}{n^2}$

3° modèle de Poisson de paramètre  $\lambda = \frac{n_a n_{\bar{b}}}{n}$

Nous retenons généralement  
l'un ou l'autre des modèles 2° et 3°  
Par exemple pour le modèle 3°:

$$\Pr [N(a \wedge \bar{b}) \geq n(a \wedge \bar{b})] = \sum_{k=\text{card}(A \cap \bar{B})}^{k=\infty} \frac{\lambda^k}{k!} \exp(-\lambda)$$

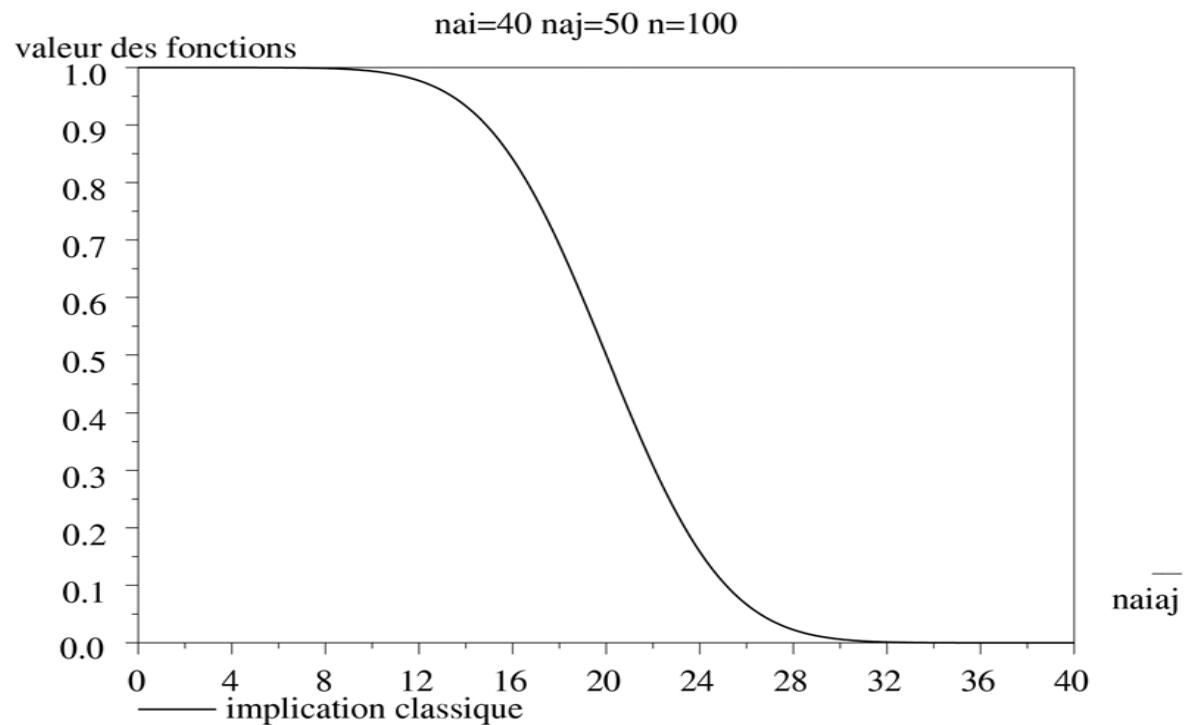
C'est la mesure de l'étonnement statistique  
donc de la quasi-implication  
pour la modélisation de Poisson

# INTENSITE D'IMPLICATION (suite)

Intensité d'implication de  $a \Rightarrow b$  :

$$\varphi(a,b) = \Pr [N(a \wedge \neg b) \geq n(a \wedge \neg b)] \text{ si } n(b) \neq n, \text{ et } \varphi(a,b) = 0 \text{ sinon}$$

Exemple de variations de  $\varphi(a_i, a_j)$  :  $n=100$ ,  $n(a_i)=40$ ,  $n(a_j)=50$  en fonction de  $n(a_i \wedge \bar{a}_j)$



# MODELE DE POISSON POUR L'INTENSITE

**L'INTENSITE D'IMPLICATION EST, PAR DEFINITION,  
LA VALEUR DE L'ETONNEMENT STATISTIQUE de POISSON**

**AINSI, dans CHIC, QUAND  $\lambda \leq 4$   
ON POSE**

$$\varphi(a, b) = \sum_{k = \text{card}(A \cap \bar{B})}^{\infty} \frac{\lambda^k}{k!} \exp(-\lambda)$$

**QUAND  $\lambda > 4$ ,  
on approxime la loi de Poisson par  
une loi normale centrée réduite en posant**

$$q = \frac{\frac{n}{a \wedge \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

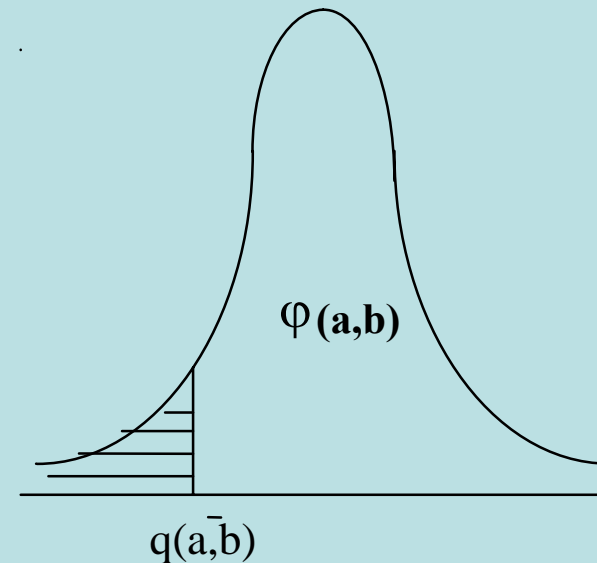
# FORME GAUSSIENNE DE L'INTENSITE

Les distributions de Poisson ou binomiale sont approximées  
en pratique par une loi normale

L'intensité d'implication s'écrit alors sous une forme gaussienne

$$\varphi(a,b) = \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

=



LA QUASI-IMPLICATION  $a \Rightarrow b$   
EST RETENUE AU NIVEAU DE CONFIANCE  $1 - \alpha$   
si  $\varphi(a,b) \geq 1 - \alpha$

# PLAN

## Introduction

§ 1 L'intensité d'implication classique (cas binaire)

**§ 2 L'implication-inclusion**

§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles

§ 4 Graphe d'implication

§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs

§ 6 Typicalité et Contribution des sujets et des variables supplémentaires

Conclusion

# IMPLICATION-INCLUSION

2 problèmes :

1° quand la population croît l'intensité d'implication  $\varphi(a,b)$  n'est pas assez discriminante.

2° l'intensité ne prend pas en compte la qualité de la contraposée.

(*Sublata causa, tollitur effectus* = la cause supprimée, l'effet disparaît)

AUSSI...

on conjugue cette intensité avec une fonction tirée de l'entropie des expériences qui mesurent une **inclusion de A dans B** où la variable a est observée ( $a=1$ ) *quel que soit b* et **l'inclusion de  $\bar{B}$  dans  $\bar{A}$**  où la variable b n'est pas observée ( $b=0$ ) *quel que soit a*.

$\Rightarrow$	b	$\bar{b}$	
a	$n_{a \wedge b}$	$n_{a \wedge \bar{b}}$	$n_a$
$\bar{a}$	$n_{\bar{a} \wedge b}$	$n_{\bar{a} \wedge \bar{b}}$	$n_{\bar{a}}$
	$n_b$	$n_{\bar{b}}$	n

Or ...

# IMPLICATION-INCLUSION (suite)

. Par **l'entropie conditionnelle  $H(b=1/a=1)$**  de b sachant a , on mesure le "déséquilibre" numérique entre les 2 cases  $n_{a \wedge \bar{b}}$  et  $n_{a \wedge b}$  du tableau de croisement de a et b en faveur de  $n_{a \wedge b}$

De la même façon, par **l'entropie conditionnelle  $H(a=0/b=0)$**  de non a sachant non b , on mesure le "déséquilibre" numérique entre les 2 cases  $n_{a \wedge \bar{b}}$  et  $n_{\bar{a} \wedge \bar{b}}$  en faveur de  $n_{\bar{a} \wedge \bar{b}}$

A partir de ces 2 fonctions, on construit 2 fonctions croissantes h1 et h2

de  $t = \frac{n_{a \wedge \bar{b}}}{n}$ , fréquence des contre-exemples qui coïncident avec ces entropies

sur  $\left[0, \frac{1}{2}\right]$  et égales à 1 sur  $\left[\frac{1}{2}, 1\right]$

# IMPLICATION-INCLUSION (suite)

Il suffit alors de considérer :

$$i(\mathbf{a},\mathbf{b}) = \left[ (1-h_1^2(\mathbf{t}))(1-h_2^2(\mathbf{t})) \right]^{\frac{1}{4}}$$

La nouvelle mesure d'intensité

$$\Psi(\mathbf{a},\mathbf{b}) = (i(\mathbf{a},\mathbf{b})\varphi(\mathbf{a},\mathbf{b}))^{\frac{1}{2}}$$

dite d'implication-inclusion, elle intègre

"**étonnement statistique**" et "**quasi-inclusion**"

Cette quantité est de nature gaussienne et permet ainsi

de retenir les valeurs les plus significatives

# PLAN

## Introduction

§ 1 L'intensité d'implication classique (cas binaire)

§ 2 L'implication-inclusion

**§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles**

§ 4 Graphe d'implication

§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs

§ 6 Typicalité et Contribution des sujets et des variables supplémentaires

Conclusion

# TYPES DE VARIABLES TRAITES

**Variables binaires** : attribut, caractère, réussite ou échec, ...

pour tout  $x \in E$ ,  $a(x) = 1$  ou  $a(x) = 0$

**Variables modales** : jugement, degré d'appartenance,...

pour tout  $x \in E$ ,  $a(x) \in [0,1]$

**Variables fréquentielles** : quantité, occurrence, ...

pour tout  $x \in E$ ,  $a(x) \in \mathbb{R}^+$

**Variables sur intervalles et variables-intervalles**:

taille, poids, échelle de notes,...

On recherche :

1° une partition optimale de sous-intervalles de valeurs

2° puis une implication optimale de réunions de ces sous-intervalles optimaux adjacents

**Un indice général** permet de mesurer  
les qualités respectives des règles entre variables  
de *même* type ou de type *différent*.

# EXEMPLE DE VARIABLE MODALE

## QUESTIONNAIRE PROPOSE A 311 ENSEIGNANTS DE CLASSES TERMINALES S, ES, LI & TE

### I Objectifs de la formation mathématique

A votre avis, quels sont les **objectifs essentiels** de la mission d'un professeur de mathématiques dans la série pour laquelle vous répondez. Pour répondre à cette question, classez par ordre préférentiel décroissant de **1** à **6** (**1** : le plus important,...) six des objectifs majeurs de cette formation en les choisissant parmi ceux des objectifs proposés ci-dessous :

**A-** acquisition de connaissances

**B-** préparation à la vie professionnelle

**C-** préparation à la vie civique et sociale

**D-** préparation aux examens, concours, au passage dans l'enseignement supérieur

**E-** développement de l'imagination et la créativité, **etc.**

Codage utilisé ::

**1** : 1 ; **2** : 0,8 ; **3** : 0,6 ; **4** : 0,4 ; **5** : 0,2 ; **6** : 0,1 ; **Vide** : 0 (par exemple, avoir placé un objectif au **rang 3** conduit à lui affecter dans l'analyse la **pondération 0,6**)

# EXEMPLE DE VARIABLE FREQUENTIELLE

Degrés de satisfaction en % dans les quartiers d'une ville (Rennes) :  
ENVIronnement, VOITure, BUS, STATionnement,  
CIRCulation, LOISirs,..

	ENVI	VOIT	BUS	STAT	CIRC	LOIS
CHMA	0,46	0,32	0,77	0,15	0,53	0,35
CENT	0,64	0,28	0,62	0,1	0,59	0,15
THAB	0,86	0,6	0,74	0,35	0,57	0,3
STHE	0,6	0,37	0,65	0,21	0,62	0,2
VEPO	0,62	0,67	0,77	0,65	0,54	0,33
SGAR	0,61	0,78	0,84	0,56	0,6	0,36
BLOS	0,63	0,71	0,84	0,67	0,58	0,33
BREQ	0,65	0,81	0,85	0,57	0,61	0,39

# EXEMPLE DE VARIABLE-INTERVALLE

## Relation taille-poids

	Taille $T_p$	Poids $P_p$	nombre de colonnes : 1+1 variables numériques
I001	180	75	-----
I002	175	70	nombre de <b>sujets</b> : 30
I003	178	77	= 1+1 variables-intervalle
I004	180	82	
I004	170	72	<u>Taille : Partitions optimales</u>
I006	198	103	T1 de 168 à 174
I007	169	68	T2 de 175 à 183
I008	172	72	T3 de 185 à 198
I009	171	76	
I010	183	80	<u>Poids : Partitions optimales</u>
I011	176	73	P1 de 67 à 72
I012	174	77	P2 de 73 à 80
.....			P3 de 82 à 103

# **PLAN**

## **Introduction**

**§ 1 L'intensité d'implication classique (cas binaire)**

**§ 2 L'implication-inclusion**

**§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles**

**§ 4 Graphe d'implication**

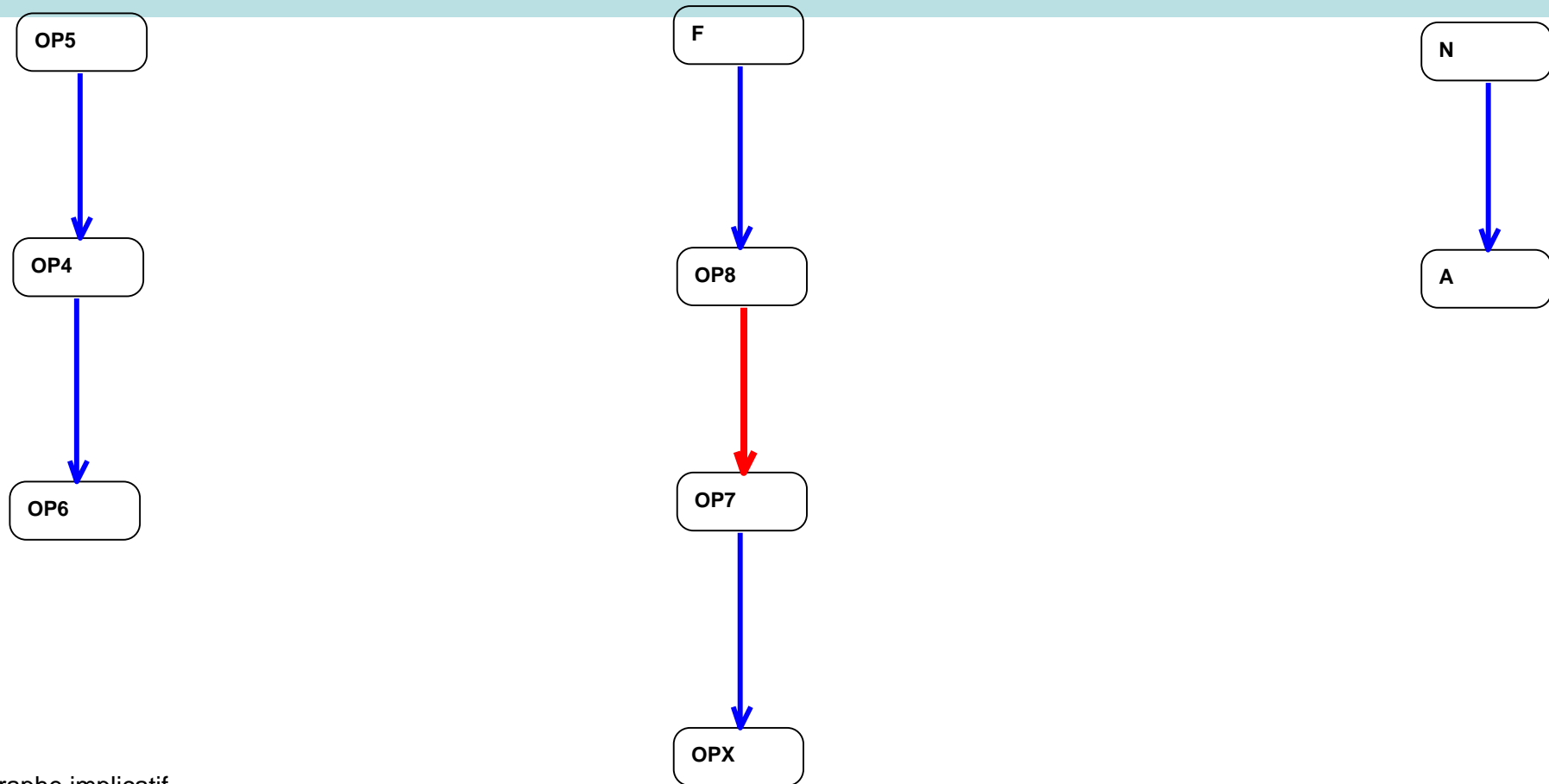
**§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs**

**§ 6 Typicalité et Contribution des sujets et des variables supplémentaires**

**Conclusion**

# EXEMPLE DE GRAPHE IMPLICATIF AU SEUIL DE 0.95

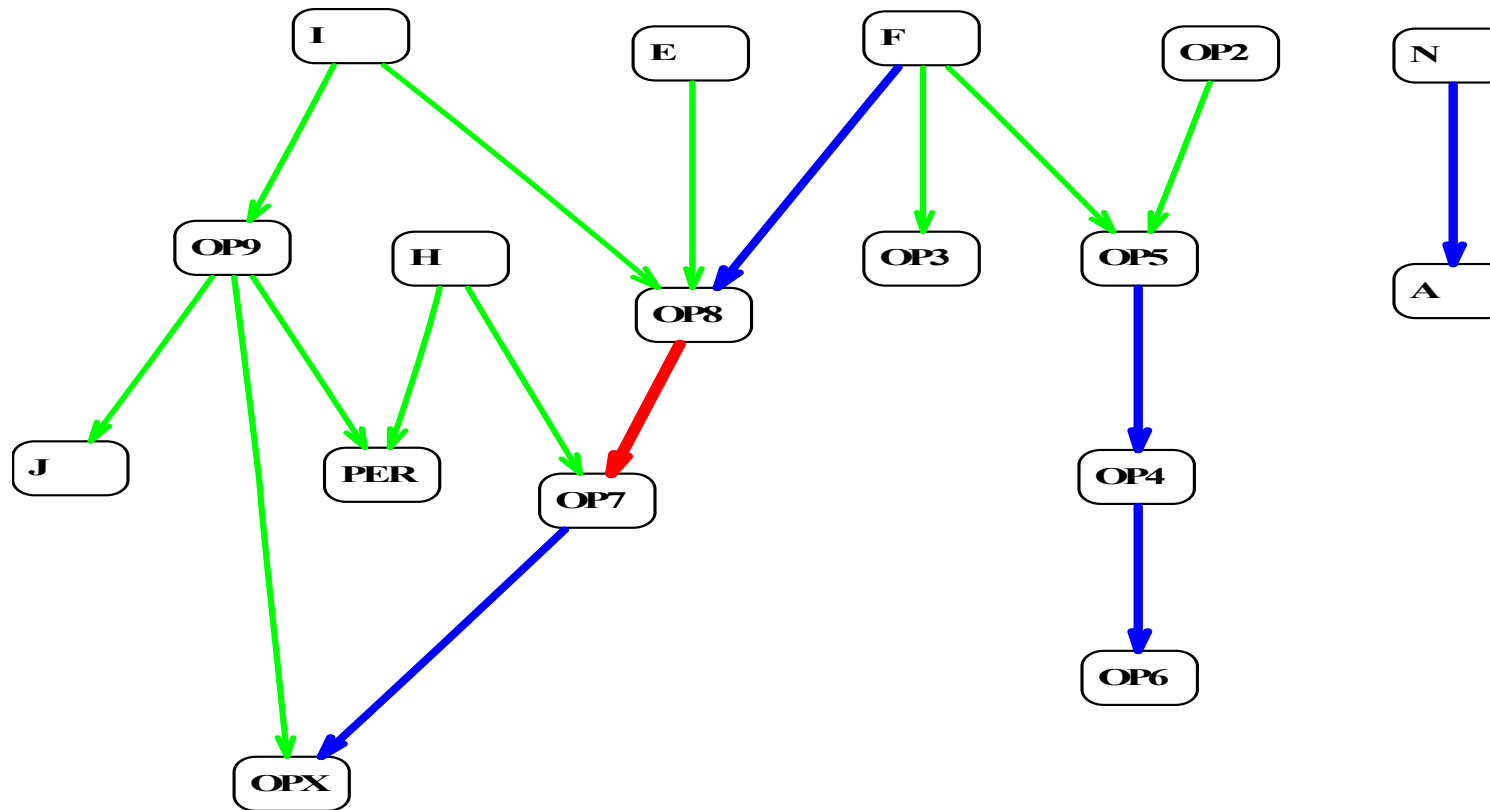
## Questionnaire enseignants-objectifs



Graphe implicatif.

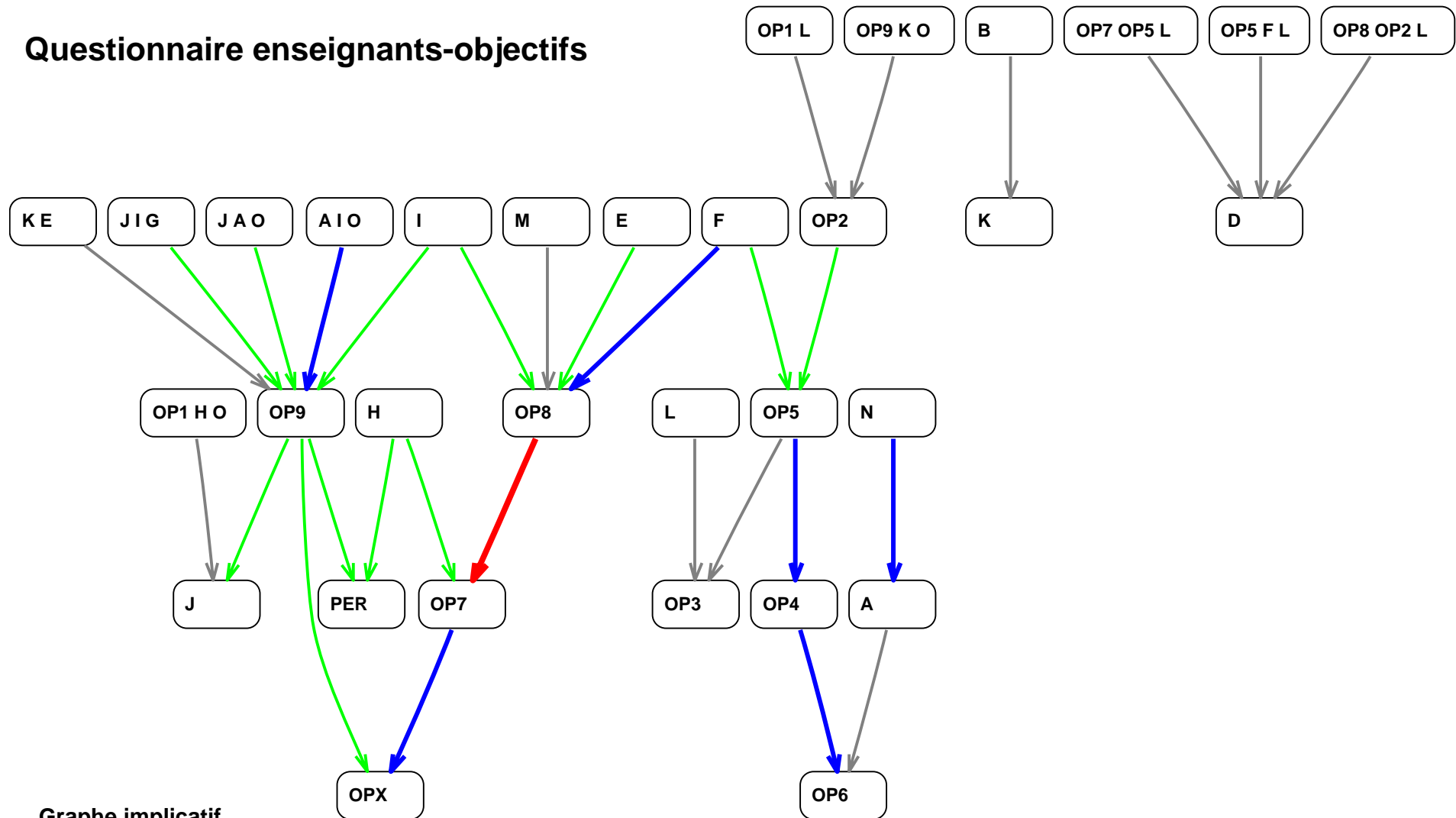
# EXEMPLE DE GRAPHE IMPLICATIF AU SEUIL DE 0.90

## Questionnaire enseignants-objectifs



# EXEMPLE DE GRAPHE IMPLICATIF A 3 PREMISSES CONJOINTES

Questionnaire enseignants-objectifs



Graphe implicatif

# EXEMPLE DE GRAPHE IMPLICATIF AU SEUIL DE 0.65

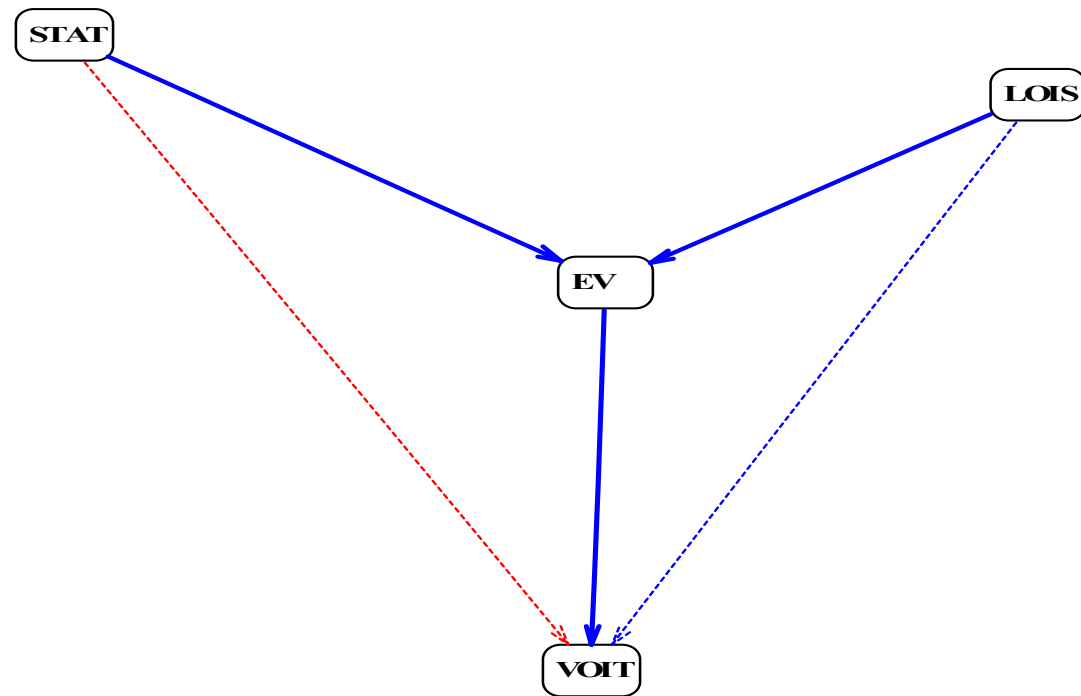
Questionnaire « degré de satisfaction en %  
dans les quartiers de Rennes »

**STAT** : qualité du stationnement

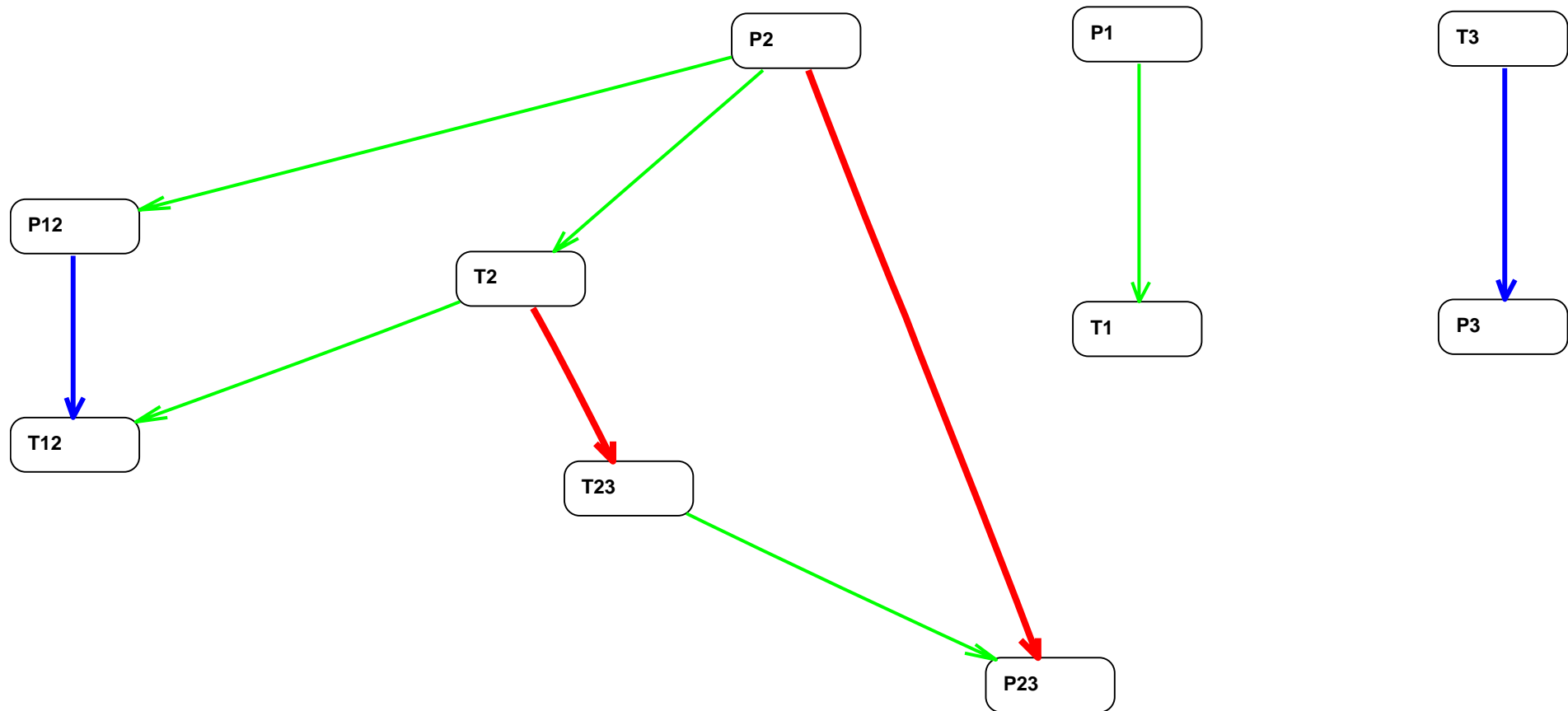
**LOIS** : présence de loisirs  
Pour les jeunes

**EV** : évolution du quartier

**VOIT** : desserte voiture



# EXEMPLE DE GRAPHE IMPLICATIF AU SEUIL DE 0.90 VARIABLE-INTERVALLE TAILLE-POIDS



# PLAN

## Introduction

§ 1 L'intensité d'implication classique (cas binaire)

§ 2 L'implication-inclusion

§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles

§ 4 Graphe d'implication

**§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs**

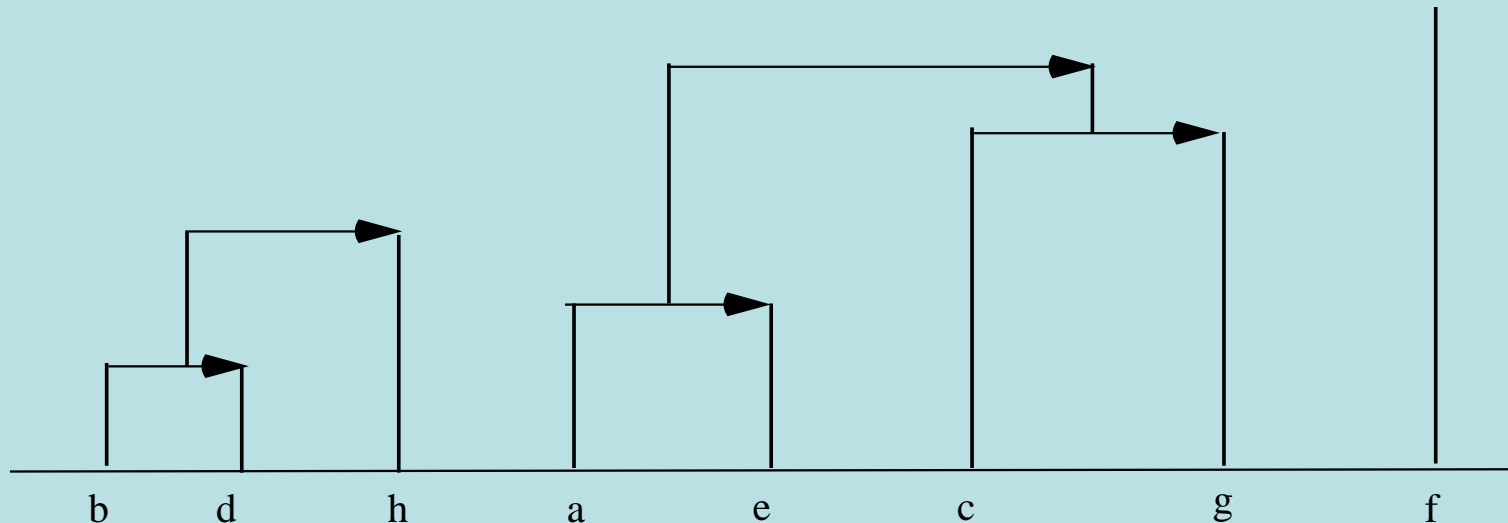
§ 6 Typicalité et Contribution des sujets et des variables supplémentaires

Conclusion

# HIERARCHIE COHESITIVE DE CLASSES

## PRINCIPE

A l'aide du **critère récursif de la cohésion**, on représente au moyen d'une **hiérarchie ascendante et indicée** (cf Gras-Kuntz, Springer, 2005) par la cohésion qui est décroissante, des "**partitions orientées**" de moins en moins fines de l'ensemble des variables.



La cohésion la plus forte est celle de la règle  $b \rightarrow d$ , suivie de  $a \rightarrow e$ , puis de  $(b \rightarrow d) \rightarrow h$ , etc.

$f$  n'entre dans aucune règle de cohésion non nulle.

L'algorithme s'arrête donc.

On dira, par exemple : *la règle  $a \rightarrow e$  quasi-implique la règle  $c \rightarrow g$*



# **PLAN**

## **Introduction**

**§ 1 L'intensité d'implication classique (cas binaire)**

**§ 2 L'implication-inclusion**

**§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles**

**§ 4 Graphe d'implication**

**§ 5 Implication entre règles et règles généralisées ;  
hiérarchie cohésitive, niveaux significatifs**

**§ 6 Typicalité et Contribution des sujets et des variables  
supplémentaires**

**Conclusion**

# TYPICALITE ET CONTRIBUTION DE SUJETS ET CATEGORIES DE SUJETS AUX GRAPHE ET HIERARCHIE IMPLICATIFS

## OBJECTIFS :

1) **mesurer** la part de "responsabilité" de sujets  
ou de catégories de sujets (variables supplémentaires)  
dans la genèse du graphe et de l'arbre,

2) **définir** des métriques sur l'ensemble des sujets

3) **distinguer** :

la **typicalité** (comportement par rapport aux règles contingentes)

et la **contribution** (comportement par rapport aux règles formelles)

(cf. Gras-David-Régnier-Guillet, *soumis EGC 2006*)

# TYPICALITE A UN CHEMIN OU A UNE CLASSE DE REGLES

## 1ère étape

### GENESE D'UN CHEMIN OU D'UNE CLASSE

(a,b) couple générique d'un chemin ( classe) C de règles constitué en dernier lieu par l'assemblage de 2 sous-chemins (classes) de variables P et Q **vérifie** :

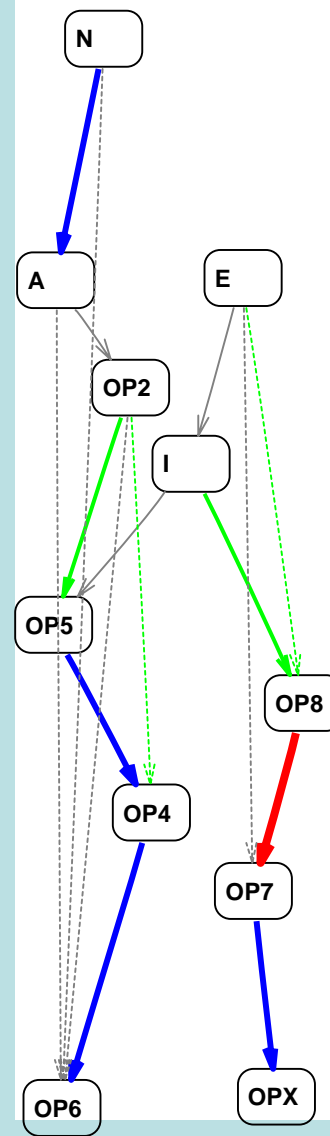
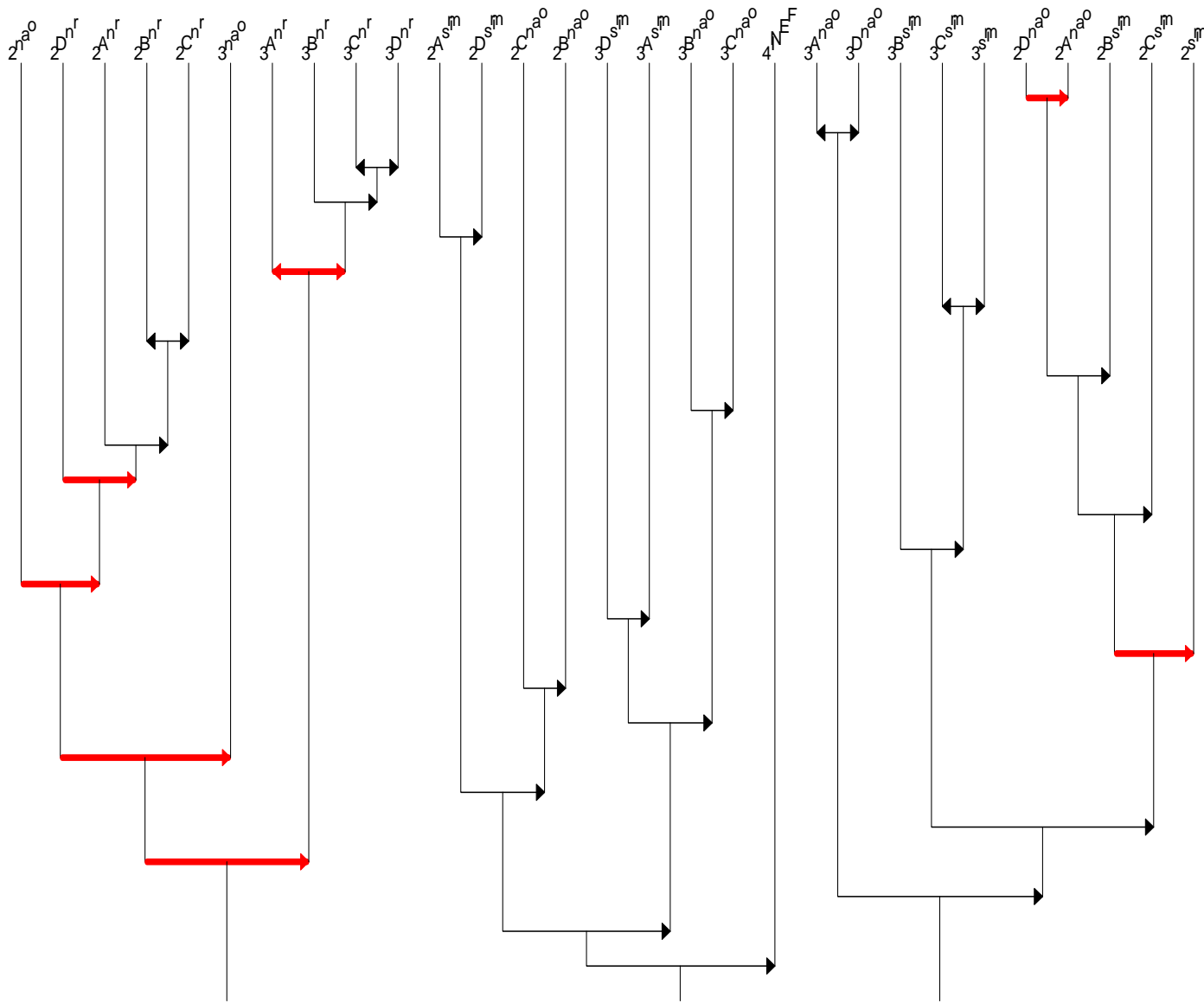
$$\forall i \in P, \forall j \in Q, \varphi(a, b) \geq \varphi(i, j)$$

$\varphi(a, b)$  , intensité maximale, est l'implication générique de C

Soit  $\varphi_1, \varphi_2, \dots, \varphi_g$  les g intensités maximales génériques apparues au cours de la constitution progressive de C

Le vecteur  $(\varphi_1, \varphi_2, \dots, \varphi_g)$  est le vecteur puissance implicative de C

# EXEMPLES DE GRAPHE ET HIERARCHIE



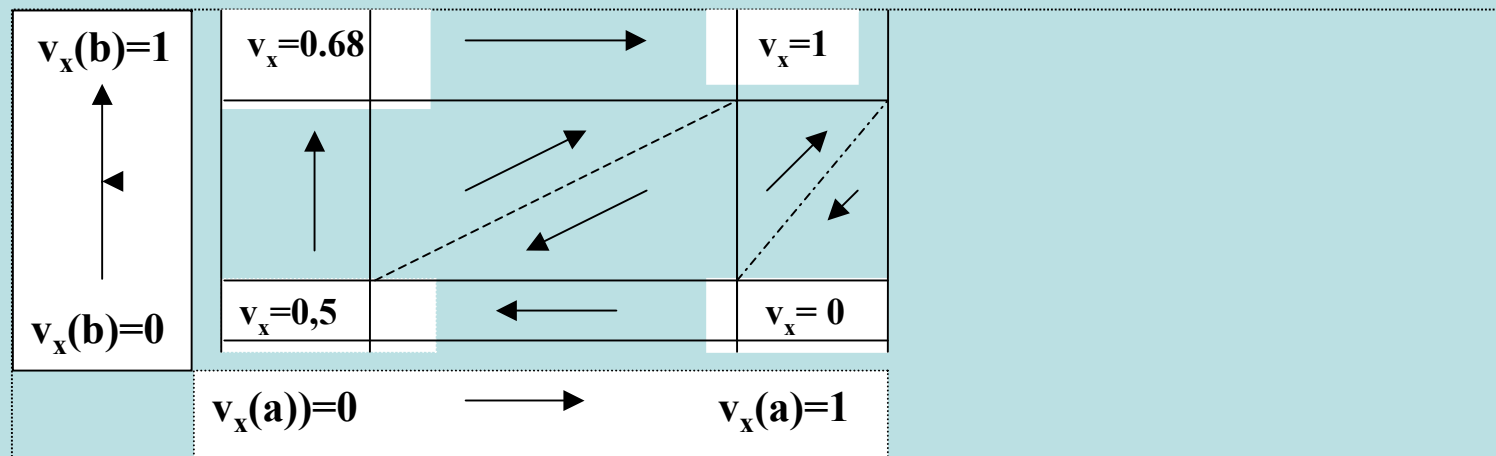
# VECTEUR PUISSANCE D'UN SUJET

Les intensités d'implication  $\varphi_{x,\cdot}$  du sujet  $x$  à chaque couple générique de  $C$  constituent un vecteur  $(\varphi_{x',1}, \varphi_{x',2}, \dots, \varphi_{x',g})$  dit **vecteur puissance implicative** de  $x$  et sont définies ainsi :

Soit  $\varphi(a,b)$  une implication générique. Si  $x$  prend les valeurs  $v_x(a)$  en  $a$  et  $v_x(b)$  en  $b$  :

- 1°  $\varphi_{x,\cdot}$  est non symétrique en  $a$  et  $b$
- 2°  $\varphi_{x,\cdot}$  est croissante avec  $v_x(a)$  et  $v_x(b)$  quand  $v_x(a) \leq v_x(b)$
- 3°  $\varphi_{x,\cdot}$  est égale à 1 quand  $v_x(a) = v_x(b) = 1$ ,  
est égale à 0,5 quand  $v_x(a) = v_x(b) = 0$   
est égale à 0 quand  $v_x(a) = 1$  et  $v_x(b) = 0$
- 4°  $\varphi_{x,\cdot}$  est décroissante quand  $v_x(b)$  étant donnée,  
 $v_x(a) > v_x(b)$  et  $v_x(a)$  croît

Dans le cas d'une variable numérique, on a construit une fonction ad hoc pour  $\varphi_{x,\cdot}$ .



## 2ème étape

# DISTANCE IMPLICATIVE

$$d(x,C) = \left[ \frac{1}{g} \sum_{i=1}^g \frac{[\varphi_{x,i} - \varphi_i]^2}{1 - \varphi_i} \right]^{\frac{1}{2}}$$

est la distance implicative de x à C

La distance entre 2 sujets :  $d(x, y) = \left[ \frac{1}{g} \sum_{i=1}^g \frac{[\varphi_{x,i} - \varphi_{y,i}]^2}{1 - \varphi_i} \right]^{\frac{1}{2}}$

définit sur E une structure d'espace métrique discret

Remarque :  $d(x,C)$  est la distance de x au sujet moyen contingent

qui aurait  $(\varphi_1, \varphi_2, \dots, \varphi_g)$  comme vecteur puissance implicative

# 3ème étape

## TYPICALITE D'UN SUJET A UN CHEMIN (CLASSE)

$$\gamma(x, C) = 1 - \frac{d(x, C)}{\max_{y \in E} d(y, C)}$$

## TYPICALITE D'UN ENSEMBLE DE SUJETS A UN CHEMIN (CLASSE)

Si G est un ensemble de sujets, sa typicalité au chemin ( classe ) C est :

$$\gamma(G, C) = \frac{1}{\text{card } G} \sum_{x \in G} \gamma(x, C)$$

# 4ème étape

## OPTIMISATION DE L'ENSEMBLE TYPIQUE A C

$\bar{\gamma}$  est le barycentre de toutes les typicalités des sujets de E

Relativement au chemin (classe) C la variance inter-classe entre l'ensemble de sujets G et son complémentaire est

$$VE = \frac{\text{cardG}}{n - \text{cardG}} (\gamma(G, C) - \bar{\gamma})^2$$

GO(C) est le **groupe optimal** de sujets :

avec son complémentaire  $\bar{GO}(C)$ , **il maximise**

**la variance inter-classe** et a une typicalité supérieure à la sienne.

# 5ème étape

Soit  $G_i$  une catégorie quelconque de sujets et  $G_i \cap GO(C)$   
les sujets de  $G_i$  qui sont dans le groupe optimal

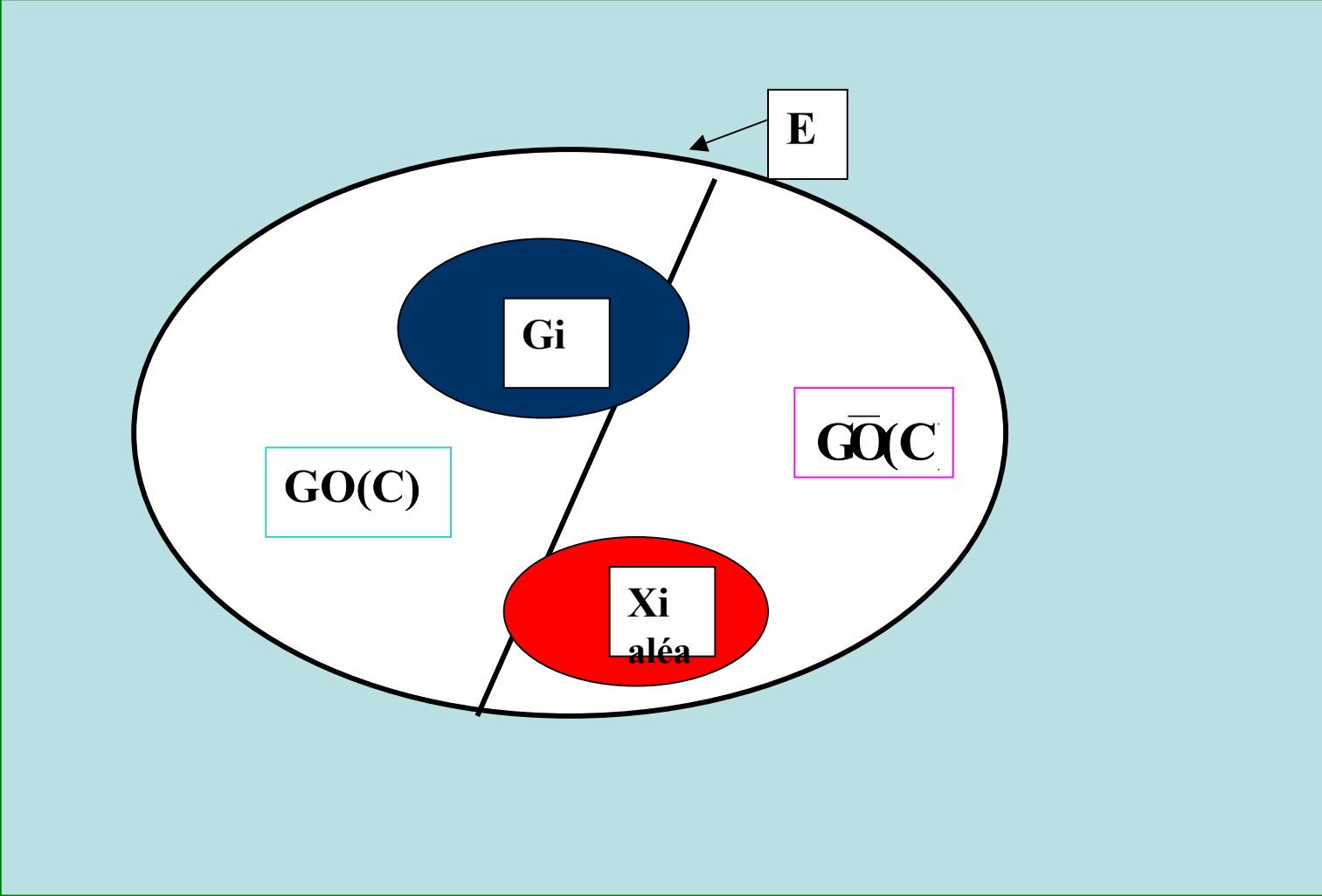
Soit  $X_i$  une partie aléatoire de  $E$  de même cardinal que  $G_i$  et  
 $Z_i = \text{card} ( X_i \cap GO(C) )$

variable aléatoire de sujets de  $X_i$  dans le groupe optimal  
de loi binomiale ( $\text{card } G_i, \text{card } GO(C)/\text{card } E$ )

La catégorie la plus typique est celle qui minimise ("étonnement statistique")  
l'ensemble des probabilités  $p_i$  telles que :

$$\forall i \quad p_i = \text{Prob} [ \text{card} ( G_i \cap GO(C) ) < Z_i ]$$

(probabilité d'avoir plus de sujets de  $G_i$  dans  $GO(C)$   
qu'il y en aurait si seul le hasard intervenait)



# 5ème étape (suite)

A chaque analyse, CHIC calcule,

pour chaque variable supplémentaire (catégorie de sujets),

- le **groupe optimal de sujets**,
- la **typicalité de chaque variable supplémentaire** à un chemin ou une classe de variables principales (ou règles)
  - et indique la **variable supplémentaire la plus typique**.

-----  
La spécificité mutuelle caractérise l'excellence privilégiée

d'un groupe ou d'un sujet relativement à une classe ou un chemin

# EXTENSION : NOTION DE CONTRIBUTION

La contribution d'un individu à un arc, à un chemin ou à une classe C est définie à partir de l'**écart** entre la **valeur prise par l'intensité d'implication** du sujet x et la **satisfaction maximale formelle** de la règle g.

Soit la **distance** :

$$d(x, C) = \left[ \frac{1}{g} \sum_{i=1}^{i=g} [1 - \varphi_{x,i}]^2 \right]^{\frac{1}{2}}$$

La **contribution de x à C** est alors :  $\Gamma(x, C) = 1 - d(x, C)$

La **contribution d'une catégorie** est définie et se calcule comme pour la typicalité.

## Typicalité à la classe : OP2,OP5,OP4,OP6 ( 2,4,8 )

### Groupe optimal :

P169 P302 P159 P213 P156 P091 P082 P072 P064 P136 P207  
P019 P044 P175 P240 P237 P205 P027 P124 P216 P094 P193  
P192 P191 P024 P141 P166 P256 P086 P214 P285 P164 P174  
P298 P018 P112 P121 P254 P238 P132 P223 P066 P225 P063  
P096 P052 P259 P036 P077 P186 P185 P101 P219 P206 P054  
P221 P158 P140 P012 P179 P067 P309 P170 P137 P059 P022  
P108 P239 P107 P109 P241 P242 P110 P084 P148 P081 P244  
P074 P060 P147 P172 P303 P180 P025 P201 P129 P311 P155  
P139 P310 P134 P277 P286 P272 P004 P095 P144 P275 P268  
P203 P057 P291 P224 P271 P028 P017 P021 P304 P270 P011  
P080 P079 P173 P120 P178 P232 P070 P143 P222 P181 P273  
P215 P199 P197 P104 P196 P195 P125 P171 P006 P262 P252  
P135 P023 P005 P016 P306 P189 P177 P297 P182 P020 P167  
P114 P236 P235 P266 P267 P085 P263 P255 P061 P071 P151  
P078 P227 P033 P283 P097 P208 P106 P034 P048 P047 P088

card GO 167 p 0.537 1-p 0.463

La variable S est typique à cette classe avec un risque de : 0.842  
intersection avec le groupe optimal 77

La variable ES est typique à cette classe avec un risque de : 0.273  
intersection avec le groupe optimal 39

La variable LI est typique à cette classe avec un risque de : 0.387  
intersection avec le groupe optimal 13

La variable TE est typique à cette classe avec un risque de : 0.264  
intersection avec le groupe optimal 38

La variable B est typique à cette classe avec un risque de : 0.899  
intersection avec le groupe optimal 7

La variable C est typique à cette classe avec un risque de : 0.369  
intersection avec le groupe optimal 12

La variable E est typique à cette classe avec un risque de : 0.471  
intersection avec le groupe optimal 6

La variable G est typique à cette classe avec un risque de : 0.941  
intersection avec le groupe optimal 3

La variable H est typique à cette classe avec un risque de : 1  
intersection avec le groupe optimal 0

La variable K est typique à cette classe avec un risque de : 0.901  
intersection avec le groupe optimal 1

La variable L est typique à cette classe avec un risque de : 0.853  
intersection avec le groupe optimal 7

La variable O est typique à cette classe avec un risque de : 0.569  
intersection avec le groupe optimal 3

**La variable la typique à cette classe est TE avec un risque de : 0.264**

# **PLAN**

## **Introduction**

### **§ 1 L'intensité d'implication classique (cas binaire)**

### **§ 2 L'implication-inclusion**

### **§ 3 Types d'autres variables traitées : modales, fréquentielles, intervalles**

### **§ 4 Graphe d'implication**

### **§ 5 Implication entre règles et règles généralisées ; hiérarchie cohésitive, niveaux significatifs**

### **§ 6 Typicalité et Contribution des sujets et des variables supplémentaires**

## **Conclusion**

# CONCLUSION ET PERSPECTIVES

- \* Le passage d'une problématique à un modèle a été illustré
  - \* la simplicité du modèle a permis sa fécondité
- \* les structures obtenues décrivent le « tout » à l'aide de « parties »

-----

- \* *stabilité des indices en fonction des paramètres*

- \* *analyse de variables vectorielles*

- \* *analyse de variables floues*

- \* *tableaux incomplets*

- *variance implicative*

- \* *arbres de classification*

.....