

# Relational Concept Analysis

An approach for classifying and mining multi-relational data

Marianne Huchard  
Montpellier University, LIRMM, France

NORTHEASTERN UNIVERSITY  
Shenyang  
2016

## Observation

- ▶ Relational datasets, multi-dimensional data
- ▶ A few approaches
  - ▶ Data warehouse (OLAP) cube (Gray et al. 2008)
  - ▶ Inductive Logic Programming (Dzeroski, 2003)
    - ▶ Propositionalization (Kuzelka, O., Zelezný, 2008)
  - ▶ Graph pattern mining (Washio and Motoda 2003, Chakrabarti and Faloutsos 2006, Liquiere and Sallantin 1998, Ganter and Kuznetsov 2001, Kuznetsov 2013)

## More specific objectives

- ▶ Extract patterns from data and relations
- ▶ Interconnected classification of entities
- ▶ Classification of extracted patterns
- ▶ Complementary point of view vs. Graph pattern mining

## Outline

- ▶ Formal Concept Analysis
- ▶ Relational Concept Analysis
- ▶ Applications

# Formal Concept Analysis

A methodology for:

- ▶ data analysis, data mining, hierarchical multi-clustering
- ▶ knowledge representation (e.g. ontology construction)
- ▶ classification, indexation (information retrieval)
- ▶ unsupervised learning (based on example description)
- ▶ supervised learning (adding classes in description)

Roots:

- ▶ lattice theory, Galois connections  
(Birkhoff, 1940; Barbut & Monjardet, 1970)
- ▶ concept lattices  
(Wille, 1982)

# Formal Concept Analysis

## Formal Context

- ▶ Simplest form: entities with characteristics

	Photo	PhysKeyboard	SquareScreen	FullHDScreen	4Gcapable	LowSAR	HighAutonomy	NFC
BlackberryPassport	x	x	x		x		x	x
BlackberryLeap	x				x		x	x
BlackberryQ5	x	x	x		x			x
iphone6plus	x			x	x		x	x
iphone6	x				x		x	x
NokiaLumia735	x				x		x	x
NokiaLumia930	x			x	x			x
SonyXperiaZ1	x			x	x		x	x
GoogleNexus5	x			x	x	x		x
AsusZE500CL	x				x			x
WikoHighway	x					x		
LGGFlex	x				x	x	x	x

SAR = specific absorption rate - NFC = Near Field Communication

# Formal context

## Formal Context $(O, A, R)$

- ▶ two finite sets  $O$  et  $A$
- ▶ binary relation  $R \subseteq O \times A$ .

## Mappings associated with $R$

- ▶ Attribute shared by an object set

$$f : \mathcal{P}(O) \rightarrow \mathcal{P}(A)$$

$$X \mapsto f(X) = \{y \in A \mid \forall x \in X, (x, y) \in R\} = X'$$

- ▶ Objects sharing an attribute set

$$g : \mathcal{P}(A) \rightarrow \mathcal{P}(O)$$

$$Y \mapsto g(Y) = \{x \in O \mid \forall y \in Y, (x, y) \in R\} = Y'$$

Alternative notation for  $f$  and  $g$ : '

# Concept

A formal **concept**  $C$  is a pair  $(E, I)$  such that  
 $f(E) = I$  (or equivalently)  $E = g(I)$

$E = \{ e \in O \mid \forall i \in I, (e, i) \in R \}$   
is the **extent** (covered objects)

$I = \{ i \in A \mid \forall e \in E, (e, i) \in R \}$   
is the **intent** (shared attributes)

# Concept

A maximal group of objects (object closed set, extent)  
sharing a maximal group of attributes (attribute closed set, intent)

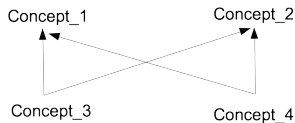
	Photo	PhysKeyboard	SquareScreen	FullHDScreen	4Gcapable	LowSAR	HighAutonomy	NFC
BlackberryPassport	x	x	x		x		x	x
BlackberryLeap	x				x		x	x
BlackberryQ5	x	x	x		x			x
iphone6plus	x			x	x		x	x
iphone6	x				x		x	x
NokiaLumia735	x				x		x	x
NokiaLumia930	x			x	x			x
SonyXperiaZ1	x			x	x		x	x
GoogleNexus5	x			x	x	x		x
AsusZE500CL	x				x			x
WikoHighway	x					x		
LGGFlex	x				x	x	x	x



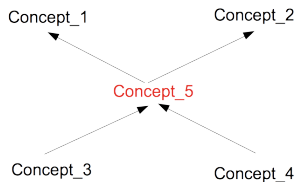
# Concept lattice

The concept set  $\mathcal{C}$  provided with:  
 $(E_1, I_1) \leq_{\mathcal{L}} (E_2, I_2) \Leftrightarrow E_1 \subseteq E_2$   
(or equivalently  $I_2 \subseteq I_1$ )  
has a lattice structure

*Every concept pair has a lower bound and an upper bound*



Partial order



Partial order with **lattice structure**

# Implication rule - exact rule

Implication rule:

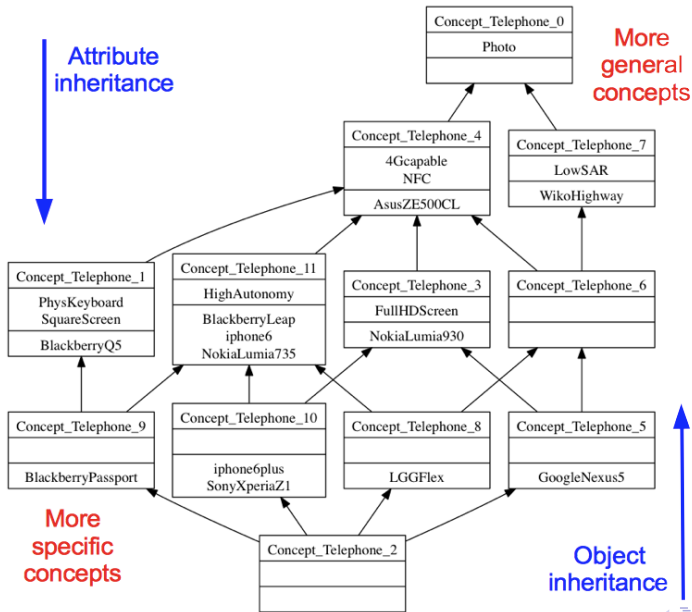
For  $A_1$  and  $A_2$  two attribute sets (itemsets)

$$A_1 \Rightarrow A_2 \text{ iff } f(A_1) \subseteq f(A_2)$$

When  $(E_1, I_1) \leq_{\mathcal{L}} (E_2, I_2)$

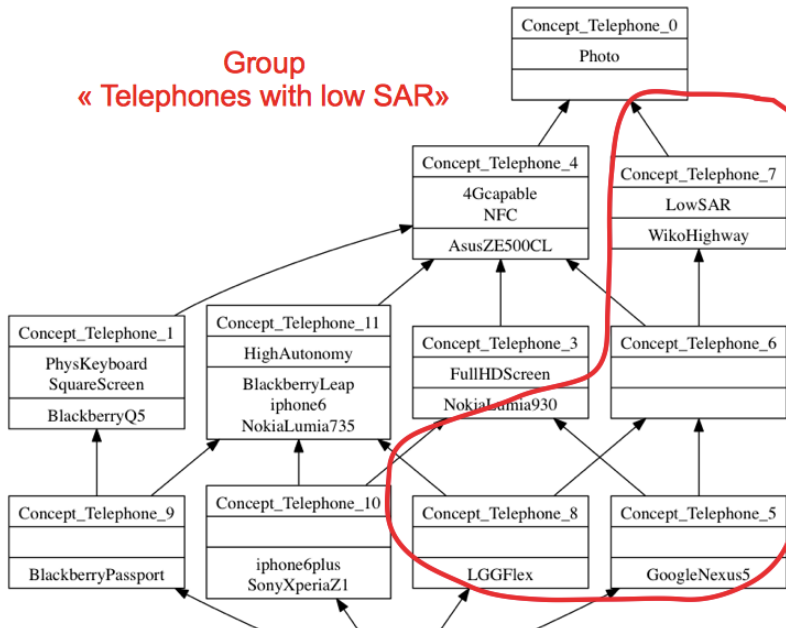
$$I_1 \Rightarrow I_2$$

# Formal Concept Analysis



# Formal Concept Analysis

Group  
« Telephones with low SAR »

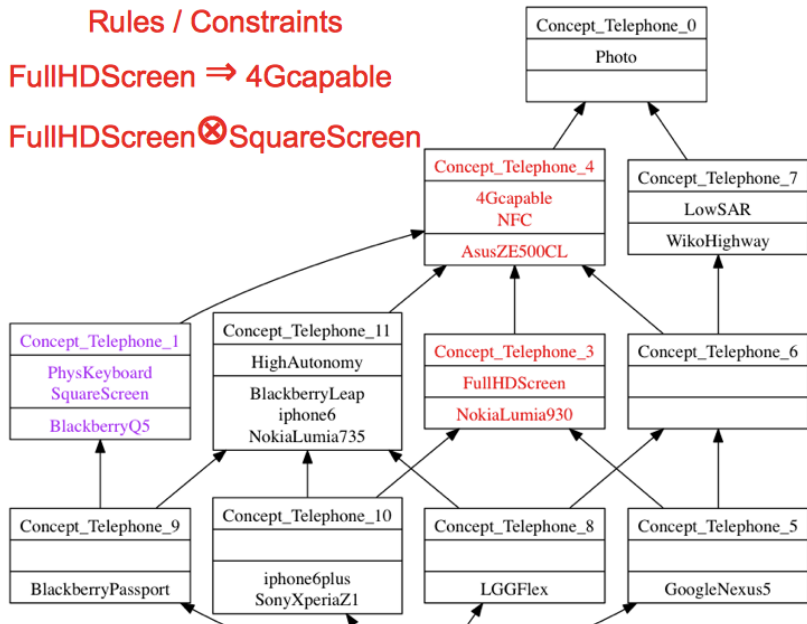


# Formal Concept Analysis

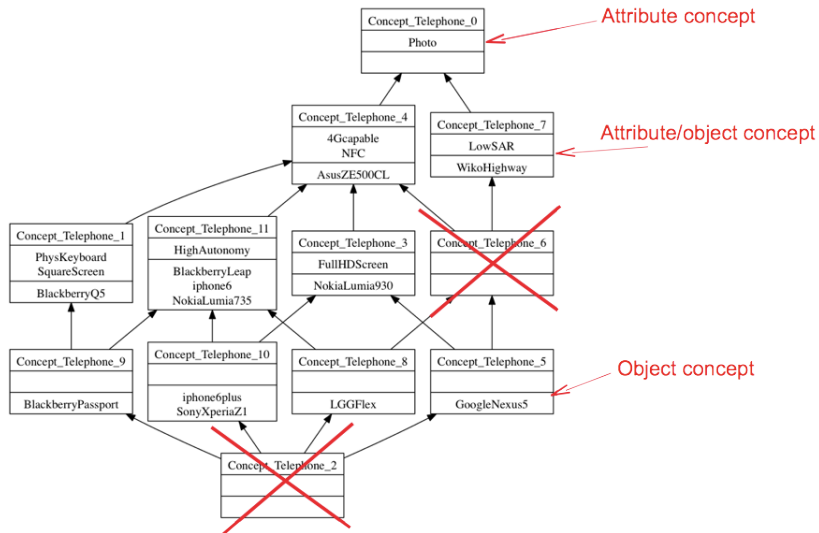
## Rules / Constraints

FullHDScreen  $\Rightarrow$  4Gcapable

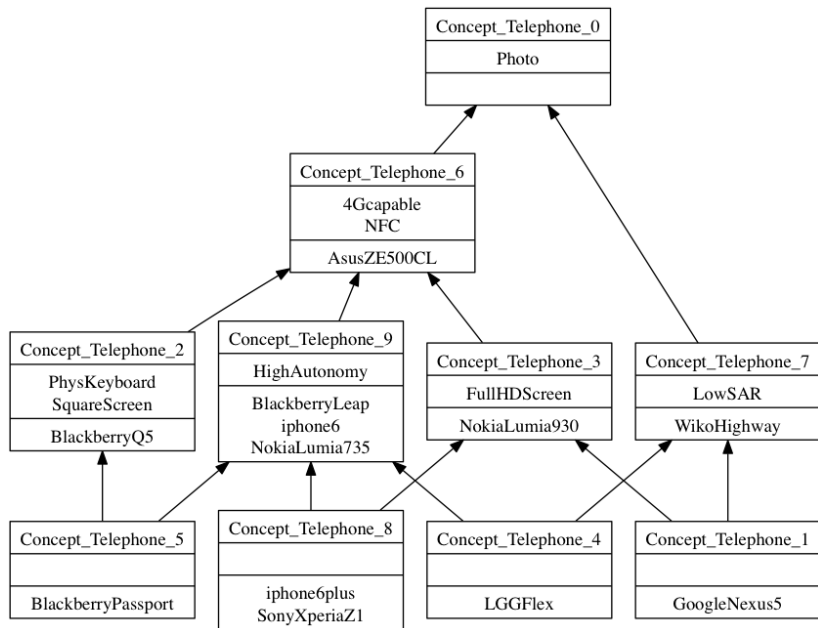
FullHDScreen  $\otimes$  SquareScreen



# AOC-poset



# AOC-poset



# Complexity of conceptual structures

## Formal context $(O, A, R)$

- ▶ two finite sets  $O$  and  $A$
- ▶ binary relation  $R \subseteq O \times A$

## Concept lattice

$$\# \text{concepts} < 2^{\min(|A|, |O|)}$$

Reached with the lattice of all subsets of  $E$ , where  $E$  is  $O$  if  $|O| = \min(|A|, |O|)$  (otherwise,  $E$  is  $A$ )

## AOC-poset

$$\# \text{concepts} < |A| + |O|$$

Reached if  $|A| = |O|$  and every attribute is shared by several distinct objects (a bipartite crown graph for ex.)



# FCA and complex data

By transforming multi-valued contexts into binary contexts (scaling) or by using a general Galois connection,

- ▶ many-valued contexts (integers, floats, terms, structures, symbolic objects, etc.) (Ganter / Wille, Polaillon, ...)
- ▶ fuzzy descriptions (Belohlavek, Yahia et al., ...)
- ▶ hierarchies on values (Godin et al., Carpineto et Romano, ...)
- ▶ logical description (Chaudron et al., Ferré et al., ...)
- ▶ graphs (Liquière, Prediger / Wille, Ganter / Kuznetsov, ...)
- ▶ pattern structures (Ganter and Kuznetsov,)
- ▶ etc.

▶ A unique classification centered on a unique object set

# Relational Concept Analysis (RCA)

- ▶ Extends the purpose of FCA for taking into account object categories and links between objects
- ▶ Main principles:
  - ▶ a relational model based on the entity-relationship model
  - ▶ integrate relations between objects as *relational* attributes
  - ▶ iterative process
- ▶ RCA provides a set of interconnected lattices
- ▶ Produced structures can be represented as ontology concepts within a knowledge representation formalism such as description logics (DLs).

Joint work with:

A. Napoli, C. Roume, M. Rouane-Hacène, P. Valtchev

# Relational Context Family (RCF)

A simple entity-relationship model to introduce RCA

## Relational Context Family

- ▶ **object-attribute contexts**
  - ▶ Pizza
  - ▶ Ingredient
- ▶ **object-object context(s)**
  - ▶ has-topping  $\subseteq$  Pizza  $\times$  Ingredient

# Relational Context Family (RCF)

A RCF  $\mathcal{F}$  is a pair  $(K, R)$  with:

- ▶  $K$  is a set of object-attribute contexts  $K_i = (O_i, A_i, I_i)$
- ▶  $R$  is a set of object-object contexts  $R_j = (O_k, O_l, I_j)$ ,
  - ▶  $(O_k, O_l)$  are the object sets of formal contexts  $(K_k, K_l) \in K^2$
  - ▶  $I_j \subseteq O_k \times O_l$
  - ▶  $K_k$  is the *source/domain context*
  - ▶  $K_l$  is the *target/range context*.
  - ▶ we may have  $K_k = K_l$ .

# RCF / object-attributes contexts

	thin	thick	calzone
Pizza			
okonomi			×
alberginia		×	
margherita	×		
languedoc	×		
four-cheeses	×		
three-cheeses	×		
frutti-di-mare	×		
quebec		×	
regina	×		
hawai		×	
lorraine	×		
kebab			×

Ingredient	fruit-vegetable	meat	fish	dairy	cereal-leguminous	veg-oil
tomato-sauce	×					
cream				×		
tomato	×					
basilic	×					
olive	×					
olive oil						×
soy	×					
mushroom	×					
eggplant	×					
onion	×					
pepper	×					
ananas	×					
mozza				×		
goat-cheese				×		
emmental				×		
fourme-ambert				×		
squid			×			
shrimp			×			
mussels			×			
ham		×				
bacon		×				
chicken		×				
maple-sirup	×					
corn					×	

# RCF / object-object context / part 1

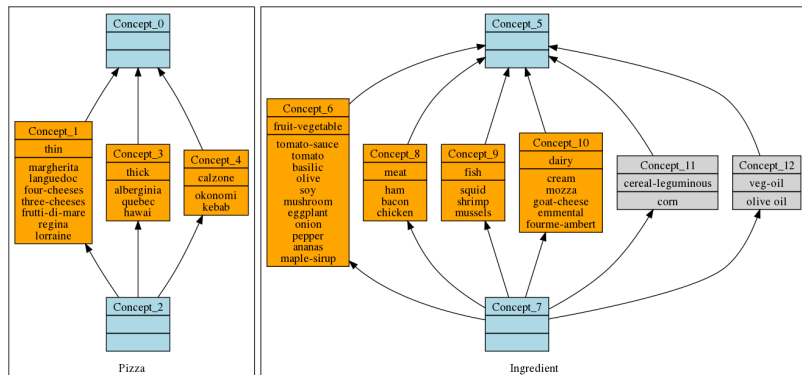
	tomato-sauce	cream	tomato	basilic	olive	olive oil	soy	mushroom	eggplant	onion	pepper	ananas
has-topping												
okonomi	x					x	x	x				
alberginia	x					x	x		x	x		
margherita	x		x	x	x	x						
languedoc	x		x	x	x	x				x	x	
four-cheeses		x										
three-cheeses		x										
frutti-di-mare	x				x	x						
quebec	x											
regina	x							x				
hawai	x											x
lorraine		x								x		
kebab	x		x		x					x		

## RCF / object-object context / part 2

	mozza	goat-cheese	emmental	fourme-ambert	squid	shrimp	mussels	ham	bacon	chicken	maple-sirup	corn
has-topping												
okonomi												
alberginia												
margherita	x											
languedoc	x											
four-cheeses	x	x	x	x								
three-cheeses	x	x	x									
frutti-di-mare	x				x	x	x					
quebec	x							x			x	x
regina	x								x			
hawai	x							x				
lorraine			x						x			
kebab			x							x		

# RCA - Initial Lattice building

At the beginning, only the object-attribute contexts are used to build the foundation of the concept lattice family





## RCA - Introducing relations as relational attributes

Then we use the current classification of ingredients by their categories and the relation to refine the classification on pizzas.

Given an object-object context  $R_j = (O_k, O_l, I_j)$ ,

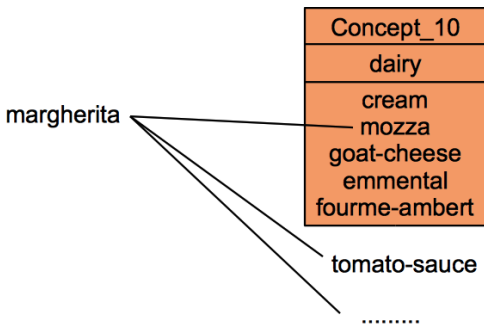
There are different relevant schemas between an object of domain  $O_k$  and concepts formed on  $O_l$ , that will be represented by relational attributes.

*E. g.* (scaling operators)

- ▶ **Existential**: an object is linked (by  $R_j$ ) to at least one object of the extent of a concept
- ▶ **Universal**: an object is linked (by  $R_j$ ) only to objects of the extent of a concept

## RCA - Existential relational attributes

**margherita** has one topping in **Concept\_10** extent: **mozza**.  
It has other links to other concept extents.



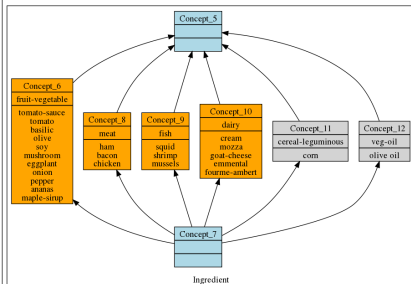
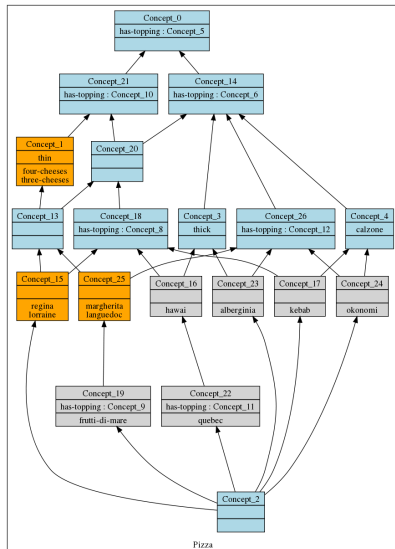
$\exists$ has-topping.Concept\_10 is assigned to **margherita**

# RCA - Existential relational attributes

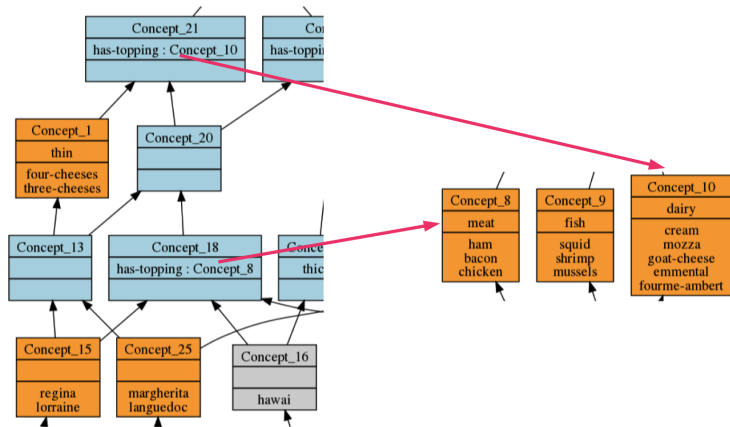
Scaled relations with domain  $O_i$  are concatenated to  $K_i$ , the object-attribute context on  $O_i$ .

Pizza	thin	thick	calzone		has-topping. Concept_7	has-topping. Concept_5	has-topping. Concept_6	has-topping. Concept_8	has-topping. Concept_9	has-topping. Concept_10	has-topping. Concept_11	has-topping. Concept_12
okonomi			x									
alberginia		x										
margherita	x											
languedoc	x			has-topping	$\exists$ has-topping. Concept_7	$\exists$ has-topping. Concept_5	$\exists$ has-topping. Concept_6	$\exists$ has-topping. Concept_8	$\exists$ has-topping. Concept_9	$\exists$ has-topping. Concept_10	$\exists$ has-topping. Concept_11	$\exists$ has-topping. Concept_12
four-cheeses	x			okonomi		x	x					x
three-cheeses	x			alberginia		x	x					x
frutti-di-mare	x			margherita		x	x			x		x
quebec		x		languedoc		x	x			x		x
regina	x			four-cheeses		x				x		
hawai		x		three-cheeses		x				x		
lorraine	x			frutti-di-mare		x	x		x	x		x
kebab			x	quebec		x	x	x		x	x	
				regina		x	x	x		x		
				hawai		x	x	x		x		
				lorraine		x	x	x		x		
				kebab		x	x	x		x		

# Relational Concept Family / exists



# Relational Concept Family / exists



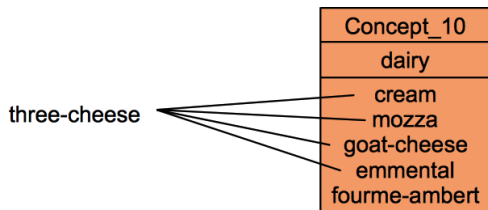
Concept\_21: pizzas with at least one topping in dairy

Concept\_18: pizzas with at least one topping in meat

have at least one meat topping  $\Rightarrow$  have at least one dairy topping

## RCA - Universal relational attributes

**three-cheese** has topping in and only in **Concept\_10** extent.



$\forall e \text{ has-topping. Concept\_10}$  is assigned to **three-cheese**

# RCA - Universal relational attributes

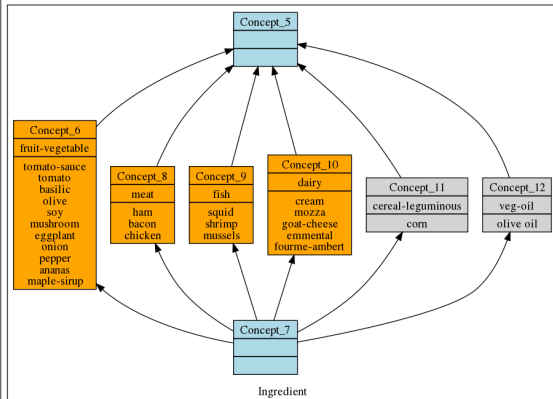
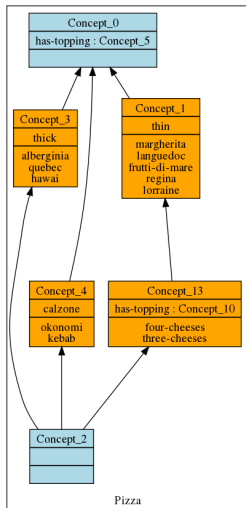
Scaled relations with domain  $O_i$  are concatenated to  $K_i$ , the object-attribute context on  $O_i$

Pizza	thin	thick	calzone
okonomi			x
alberginia		x	
margherita	x		
languedoc	x		
four-cheeses	x		
three-cheeses	x		
frutti-di-mare	x		
quebec		x	
regina	x		
hawai		x	
lorraine	x		
kebab			x

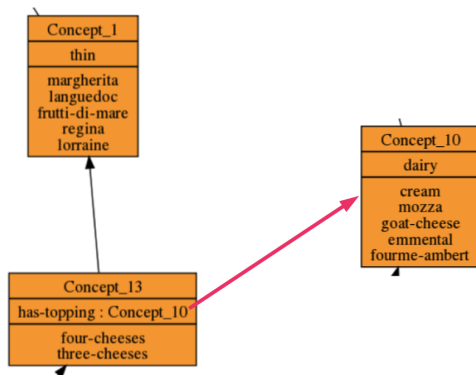
has-topping	$\forall \exists \text{has-topping. Concept\_7}$	$\forall \exists \text{has-topping. Concept\_5}$	$\forall \exists \text{has-topping. Concept\_6}$	$\forall \exists \text{has-topping. Concept\_8}$	$\forall \exists \text{has-topping. Concept\_9}$	$\forall \exists \text{has-topping. Concept\_10}$	$\forall \exists \text{has-topping. Concept\_11}$	$\forall \exists \text{has-topping. Concept\_12}$
okonomi		x						
alberginia		x						
margherita		x						
languedoc		x						
four-cheeses		x				x		
three-cheeses		x				x		
frutti-di-mare		x						
quebec		x						
regina		x						
hawai		x						
lorraine		x						
kebab		x						

# Relational Concept Family / forall





# Relational Concept Family / forall



Concept\_13: pizzas with only dairy topping

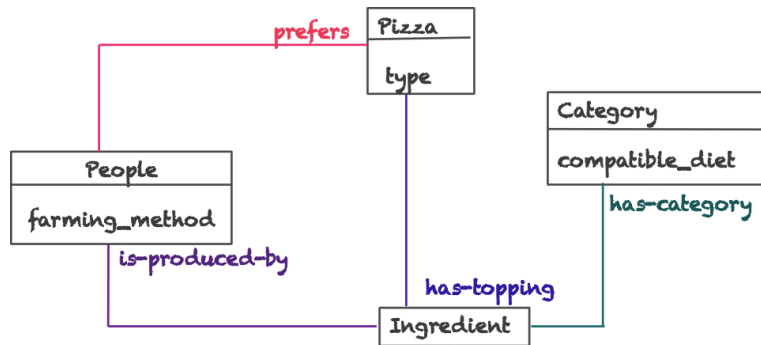
Concept\_1: thin pizzas

have only dairy topping  $\Rightarrow$  thin

## A variety of scaling operators to tune the results

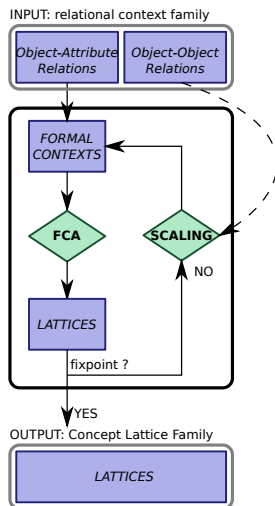
Operator	Attribute form	Condition
Universal (narrow)	$\forall r.c$	$r(o) \subseteq Ext(c)$
Universal strict	$\forall \exists r.c$	$r(o) \subseteq Ext(c)$ and $r(o) \neq \emptyset$
Universal-percent	$\forall \exists \_ \geq n\% r.c$	$ r(o) \cap Extent(C)  \geq n r(o) /100$
Covers	$\supseteq r.c$	$r(o) \supseteq Ext(c)$
Covers-percent	$\supseteq \_ \geq n\% r.c$	$ r(o) \cap Extent(C)  \geq n Extent(C) /100$
Existential (wide)	$\exists r.c$	$r(o) \cap Ext(c) \neq \emptyset$
Universal strict	$\forall \exists r.c$	$r(o) \subseteq Ext(c)$ and $r(o) \neq \emptyset$
Qualif. card. restriction	$\geq n r.c$	$r(o) \subseteq Ext(c)$ and $ r(o)  \geq n$
Card. restriction	$\geq n r.T_{\mathcal{L}}$	$ r(o)  \geq n$

# General Entity-Relationship diagram

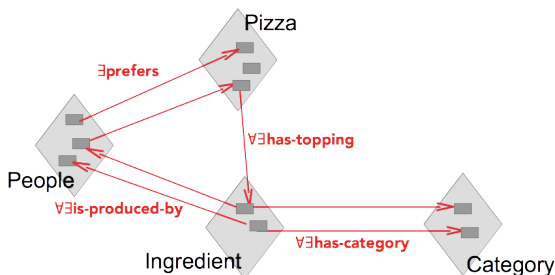


- ▶ General ER diagram may present cycle/circuits between classes/objects
- ▶ Scaling operators are assigned to relations:  $\exists$  prefers  $\forall\exists$  has-topping  $\forall\exists$  has-category  $\forall\exists$  is-produced-by

# RCA schema



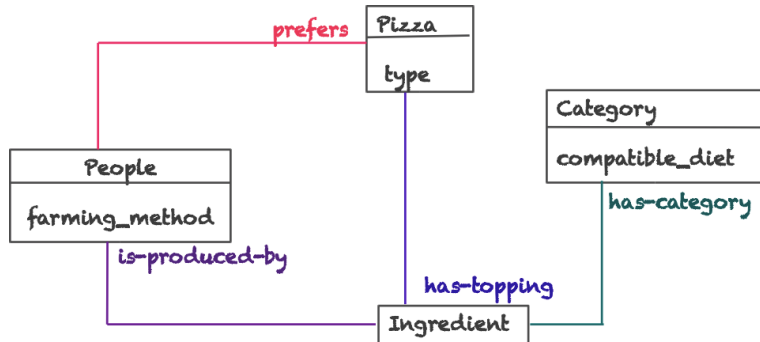
# Interconnected lattices



Example of possible learned knowledge

- ▶  $\forall \exists \text{has-category.Vegetable} \Leftrightarrow \forall \exists \text{is-produced-by.Organic farmers}$
- ▶ A subgroup of organic farmers prefer at least one pizza with only vegan topping ingredients and produced only by organic farmers

# Analysis of pizza data



$\exists$  prefers  $\forall \exists$  has-topping  $\forall \exists$  has-category  $\forall \exists$  is-produced-by

# Analysis of pizza data - object-attribute contexts

Pizza	thin	thick	calzone
forest		×	
occitane			×
three-cheese	×		
four-cheese	×		
lorraine	×		
arctic		×	

People	organic-farmer	conventional-farmer
Amedeo	×	
Amine	×	
Cyril		×
Marianne	×	
Petko		×

Ingredient
tomato-sauce
cream
onion
bacon
salmon
soy-cream
mozza
goat-cheese
emmental
fourme-ambert
eggplant
mushroom

Category	mediterranean	vegan	vegetarian
fruit-vegetable	×	×	×
meat			
fish	×		
dairy	×		×

# Analysis of pizza data - object-object contexts

prefers	forest	occitane	three-cheese	four-cheese	lorraine	arctic
Amedeo	×					
Amine		×				
Cyril				×	×	
Marianne			×			×
Petko						×

has-topping	tomato-sauce	cream	onion	bacon	salmon	soy-cream	mozza	goat-cheese	emmental	fourme-ambert	eggplant	mushroom
forest						×						×
occitane	×		×								×	
three-cheese	×						×	×	×			
four-cheese	×	×					×	×	×	×		
lorraine		×	×	×			×					
arctic	×	×			×		×					

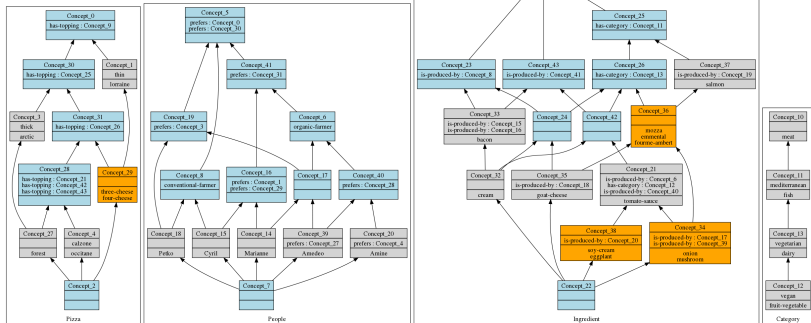


# Analysis of pizza data - object-object contexts

is-produced-by	Amedeo	Amine	Cyril	Marianne	Petko
tomato-sauce	×	×			
cream			×		
onion	×				
bacon			×		
salmon				×	×
soy-cream		×			
mozza				×	×
goat-cheese					×
emmental				×	×
fourme-ambert				×	×
eggplant		×			
mushroom	×				

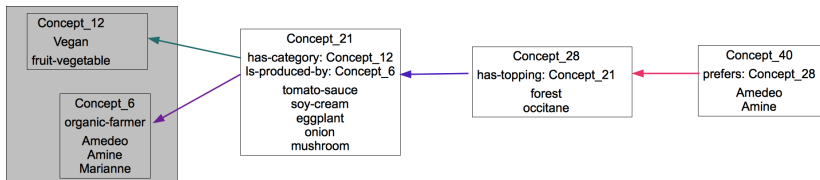
has-category	fruit-vegetable	meat	fish	dairy
tomato-sauce	×			
cream				×
onion	×			
bacon		×		
salmon			×	
soy-cream	×			
mozza				×
goat-cheese				×
emmental				×
fourme-ambert				×
eggplant	×			
mushroom	×			

# Concept lattice family



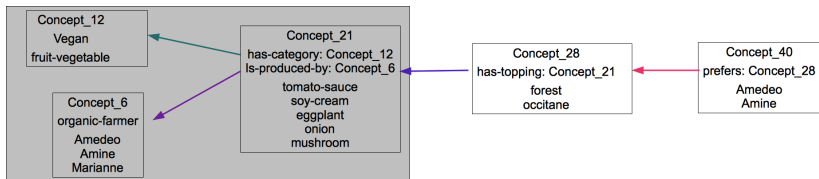
$\exists$  prefers  $\forall \exists$  has-topping  $\forall \exists$  has-category  $\forall \exists$  is-produced-by

# Concept lattice family



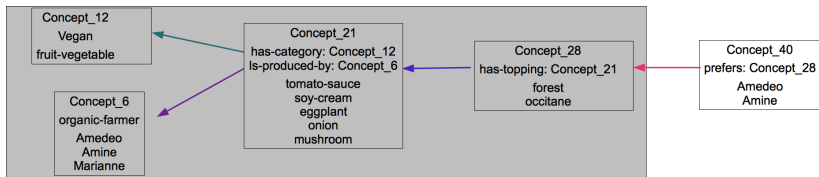
Step 0

# Concept lattice family



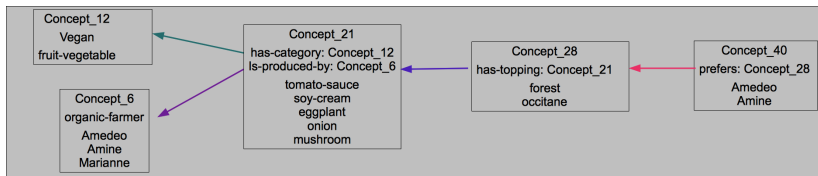
Step 1

# Concept lattice family



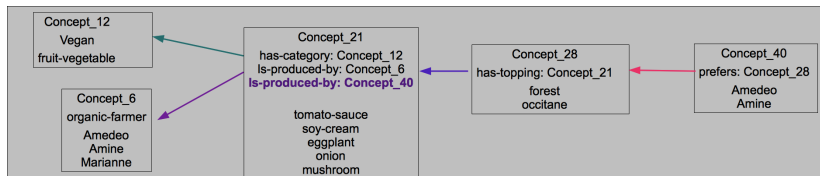
Step 2

# Concept lattice family



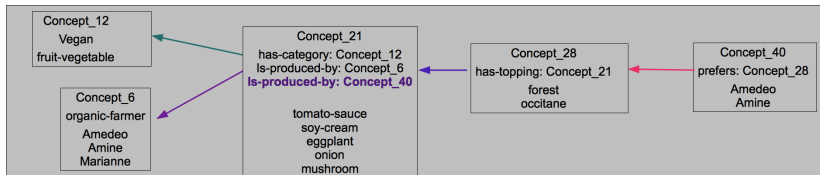
Step 3

# Concept lattice family



Step 4

# Concept lattice family



- ▶ People:  $\exists \text{prefers. Concept\_28} \Rightarrow$  organic farmer
- ▶ Ingredient:  $\forall \exists \text{has-category. Concept\_12} \Leftrightarrow \forall \exists \text{is-produced-by. Concept\_6}$  (organic farmers)
- ▶ Amedeo/Amine prefer at least one pizza with only vegan topping ingredients and produced only by organic farmers



# Applications to Knowledge Representation and Data Mining

- ▶ Ontology construction (R. Bendaoud, M. Rouane Hacene, Y. Toussaint, B. Delecroix, A. Napoli)
- ▶ Ontology restructuring (M. Rouane-Hacene, R. Nkambou and P. Valtchev)
- ▶ Discovering hidden user profiles in a semantic actors-activities network (Z. Azmeh, I. Mirbel)
- ▶ Information retrieval in legal document collections (queries) (N. Mimouni, A. Nazarenko, S. Salotti)

# FRESQUEAU project (ANR11\_MONU14)

Joint work with: A. Braud, X. Dolques, C. Grac, F. Le Ber, C. Nica  
<http://engees-fresqueau.unistra.fr>

- ▶ Develop new methods to study, compare and exploit the whole set of available parameters describing the state of watercourses
- ▶ Extraction of implication rules (with premise of size 1)  
*Presence of taxons of size from 0 to 2 cm and with a lifetime of 1 month implies presence of SO4*
- ▶ 4 OA-contexts (49 sites, 197 macro-invertebrates, 27 Physico-Chemical parameters, 18 life traits (116 modalities)), 5+6+3 OO-contexts
- ▶ variants on scaling operators
- ▶ Lattices: > 10 000 concepts
- ▶ AOC-posets: from ~600 to ~1500 concepts, ~130 to ~300 rules

## Reengineering of existing software, by building new software artefacts

- ▶ UML class diagram refactoring (M. Dao, M. Huchard, M. Rouane Hacene, C. Roume, P. Valtchev, G. Arévalo, J.-R. Falleri, C. Nebut)
- ▶ UML Use case diagram refactoring (X. Dolques, M. Huchard, C. Nebut, P. Reitz)
- ▶ Blob design defect correction (N. Moha, M. Rouane Hacene, P. Valtchev, Y.-G. Guéhéneuc)
- ▶ Extracting architectures in object-oriented software (A.-E. El Hamdouni, A. Seriai, M. Huchard)

# Applications to Software engineering

## Learning from model transformation examples, and inferring transformation rules

- ▶ Learning model transformation patterns in MDE (H. Saada, X. Dolques, M. Huchard, C. Nebut, H. A. Sahraoui)

## Classification of software artefacts

- ▶ Classification of components (G. Arévalo, M. Huchard, C. Urtado, S. Vauttier)
- ▶ Classification of web services (Z. Azmeh, M. Driss, F. Hamoui, M. Huchard, N. Moha, C. Tibermacine)

## Software analysis

- ▶ Analysis of the evolution of class diagrams (A. Osman-Guédi, A. Miralles, B. Amar, M. Huchard, T. Libourel and C. Nebut)

# Conclusion

- ▶ RCA: an opportunity for analyzing more deeply datasets composed of objects and relations
- ▶ Can be mixed with other FCA extension (e.g. numerical data)
- ▶ Exploratory RCA allows us step-by-step analysis, considering a subset of the dataset and changing structures (lattices, AOC-posets, iceberg)

## Tools

- ▶ Galicia: <http://galicia.sourceforge.net/>
- ▶ eRCA: <http://code.google.com/p/erca/>
- ▶ RCAexplore: <http://dolques.free.fr/rcaexplore/>

- ▶ Ongoing work on temporal data in the context of Fresqueau project
- ▶ Guided exploration
  - ▶ Local data exploration by incremental algorithms
  - ▶ Develop querying approach, or metrics
  - ▶ Analyzing the specialization relations between lattices built with different scaling operators
- ▶ Visualization

# Thank you / Questions?

Marianne Huchard, Mohamed Rouane Hacene, Cyril Roume, Petko Valtchev: **Relational concept discovery in structured datasets** Ann. Math. Artif. Intell. 49(1-4): 39-76 (2007)

Mohamed Rouane Hacene, Marianne Huchard, Amedeo Napoli, Petko Valtchev: **Relational concept analysis: mining concept lattices from multi-relational data** Ann. Math. Artif. Intell. 67(1): 81-108 (2013)