Combining Dependent Annotations for Relational Algebra

Egor V. Kostylev, Peter Buneman

University of Edinburgh

Theory and Practice of Provenance, 2011

Semiring Model

- Domain of annotations for positive relational algebra (SPJU) is expected to be a semiring [Green, et al.]
- What to do if we need to annotate a database with 2 domains R_1 and R_2?
- Simple answer: **the set of pairs** $R_1 \times R_2$.
- Does it always work?

Semiring Model

- Domain of annotations for positive relational algebra (SPJU) is expected to be a semiring [Green, et al.]
- What to do if we need to annotate a database with 2 domains R_1 and R_2?
- Simple answer: **the set of pairs** $R_1 \times R_2$.
- Does it always work?

Semiring Model

- Domain of annotations for positive relational algebra (SPJU) is expected to be a semiring [Green, et al.]
- What to do if we need to annotate a database with 2 domains R_1 and R_2?
- Simple answer: **the set of pairs** $R_1 \times R_2$.
- Does it always work?

Example

Exports:

CName	Goods	Time	Customers
Greece	Food	2004-2008	UK, Germany
Greece	Textile	2007-2010	Germany, Italy, Cyprus

Time – sets of **years** with $\cup$ and $\cap$ as operations
Customers – sets of **countries** with $\cup$ and $\cap$ as operations

Q $= \pi_{\text{CName}}(\text{Exports})$:

CName	Time	Customers
Greece	2004-2010	UK, Germany, Italy, Cyprus

Is it the answer we expect?

Example

Exports:

CName	Goods	Time	Customers
Greece	Food	2004-2008	UK, Germany
Greece	Textile	2007-2010	Germany, Italy, Cyprus

Time – sets of **years** with $\cup$ and $\cap$ as operations
Customers – sets of **countries** with $\cup$ and $\cap$ as operations

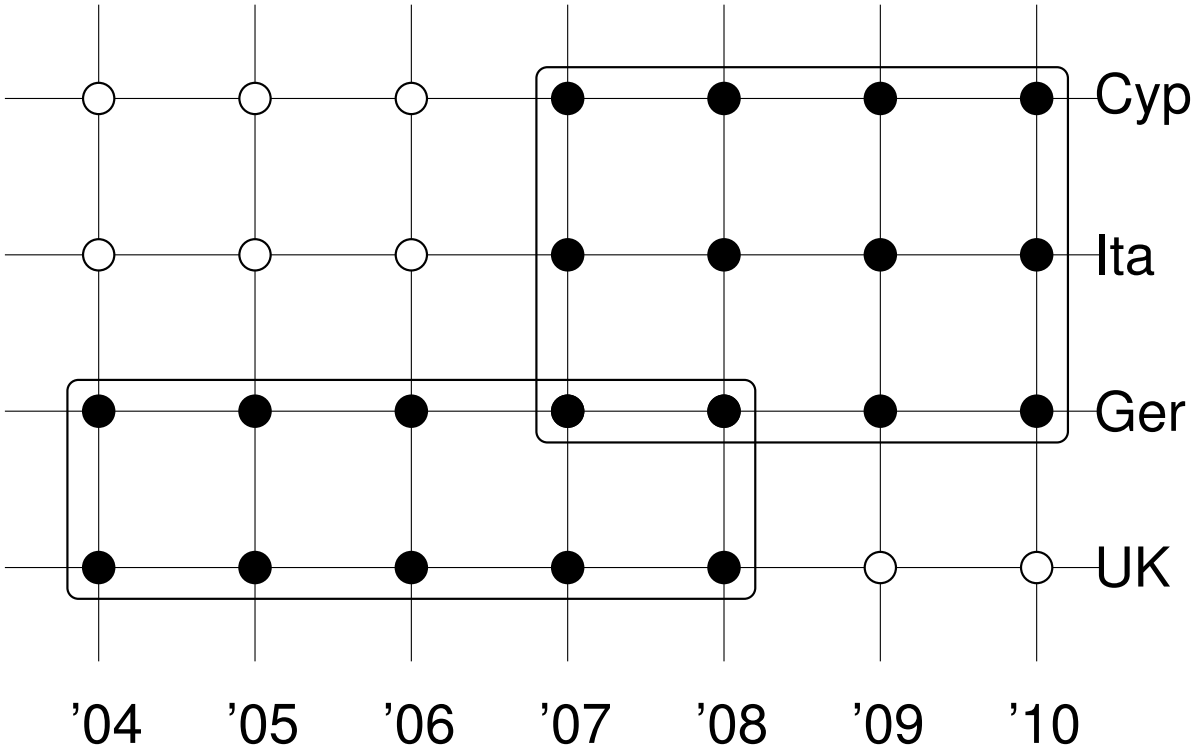
Q $= \pi_{\text{CName}}(\text{Exports})$:

CName	Time	Customers
Greece	2004-2010	UK, Germany, Italy, Cyprus

Is it the answer we expect?

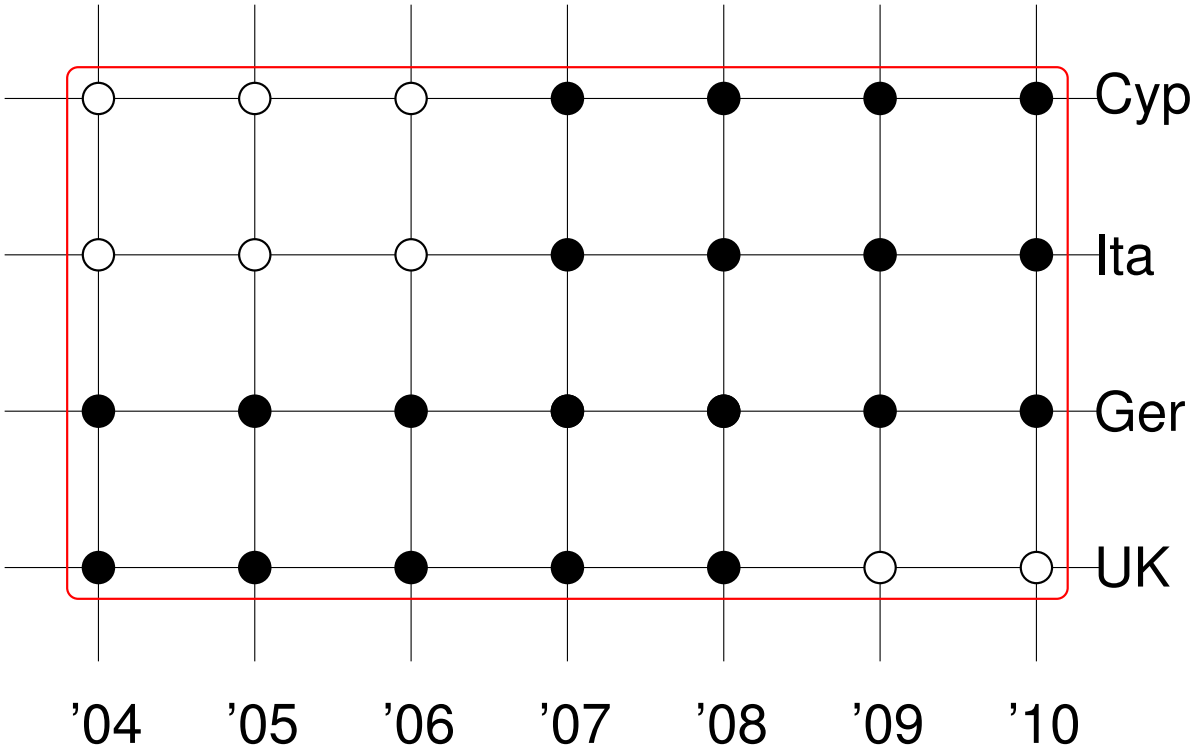
Graphical representation

([2004-2008], {UK, Germany})
([2007-2010], {Germany, Italy, Cyprus}):



Graphical representation

([2004-2010], {UK, Germany, Italy, Cyprus})



Combined domain of dependent annotations

It is impossible to represent the desired set of dots by a single pair of elements from the combining domains.

Combined annotation – a set of pairs from $R_1 \times R_2$.

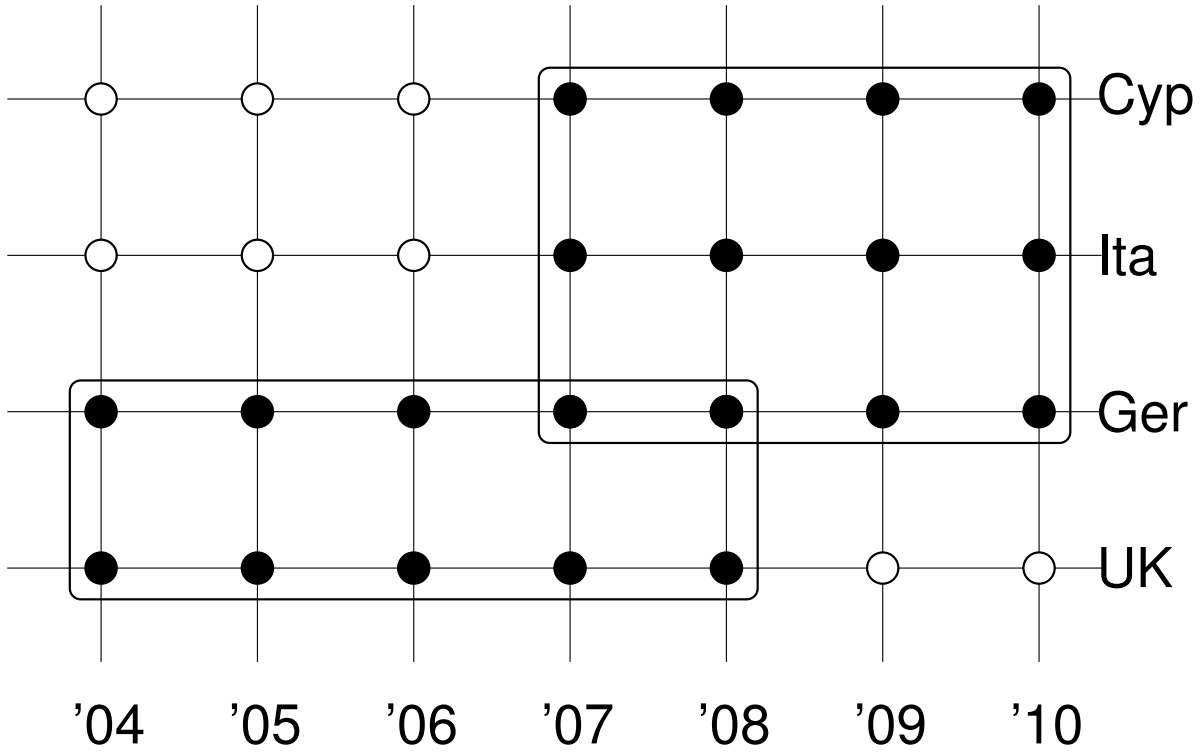
Combined domain of dependent annotations

It is impossible to represent the desired set of dots by a single pair of elements from the combining domains.

Combined annotation – a set of pairs from $R_1 \times R_2$.

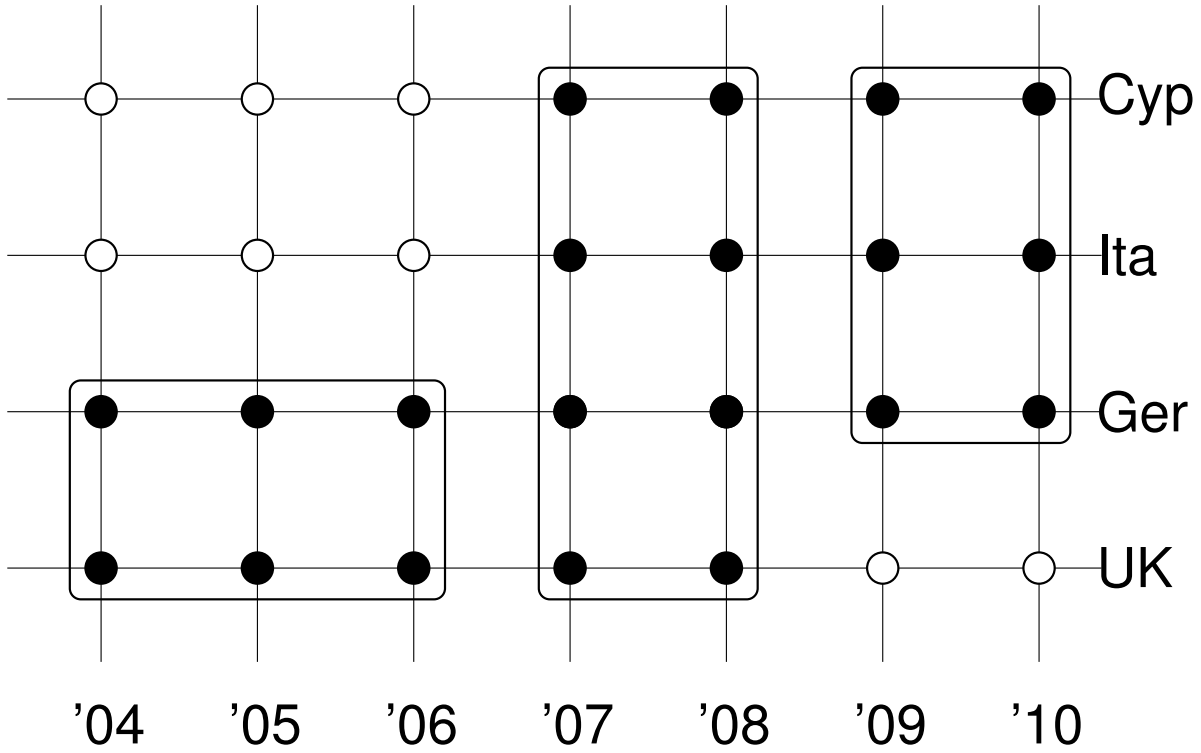
Example: Combined annotation

λ_1 = {([2004-2008], {UK, Germany})
([2007-2010], {Germany, Italy, Cyprus})}:



Example: Combined annotation

λ_2 = {([2004-2006], {UK, Germany})
([2007-2008], {UK, Ger, Italy, Cyprus})}:
([2009-2010], {Germany, Italy, Cyprus})}:



Semiring of Combined Annotations

- define an **equivalence** in combined annotations
- define a **semiring** of equivalence classes of combined annotations
- define a **normal form** for equivalence classes
- design an **algorithm** to compute normal forms

Do it carefully to make it work for (almost) all semirings (no difference, idempotence, etc.)