Formal Concept Analysis

III Knowledge Discovery

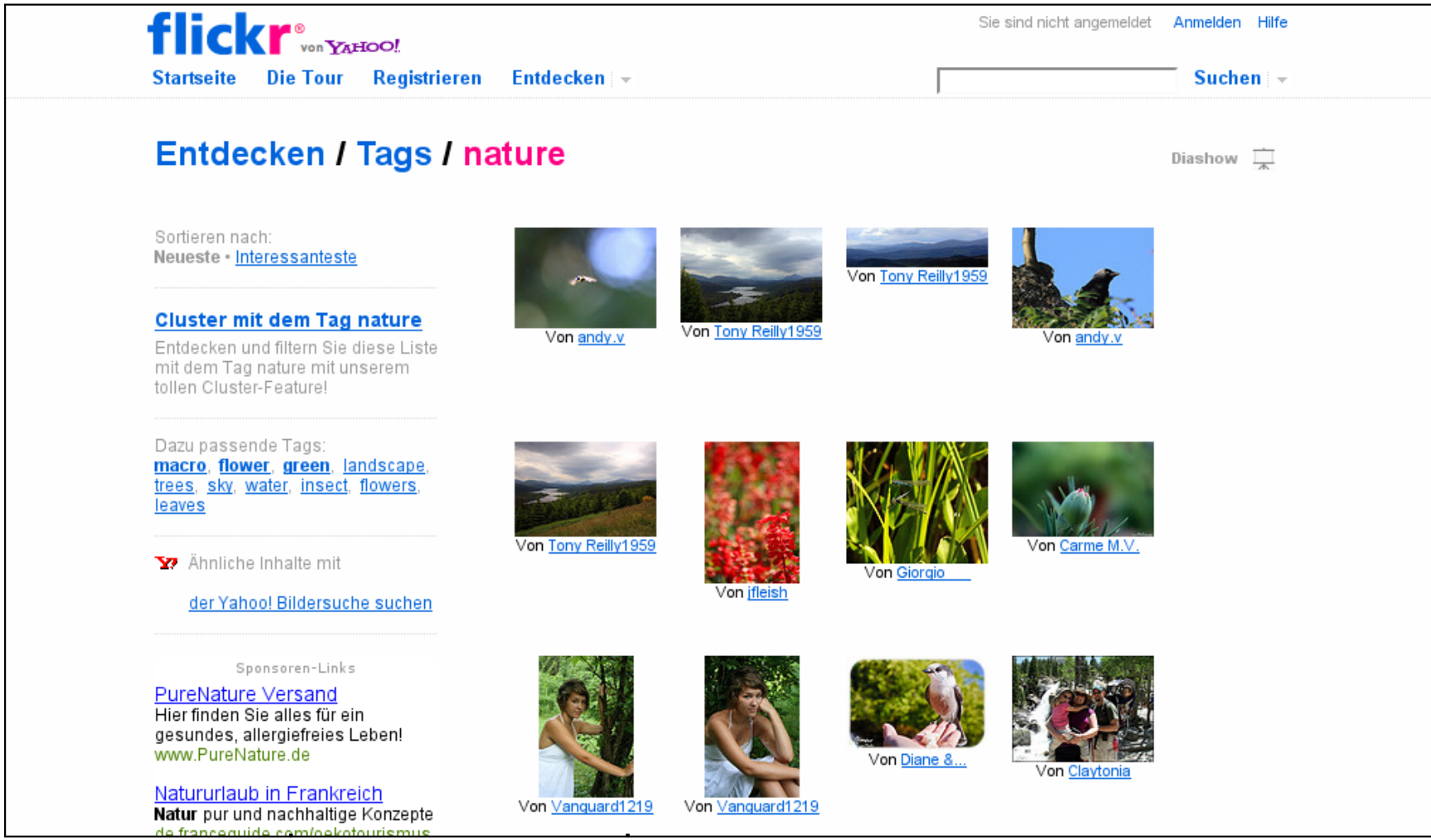
Sebastian Rudolph

Computational Logic Group
Technische Universität Dresden

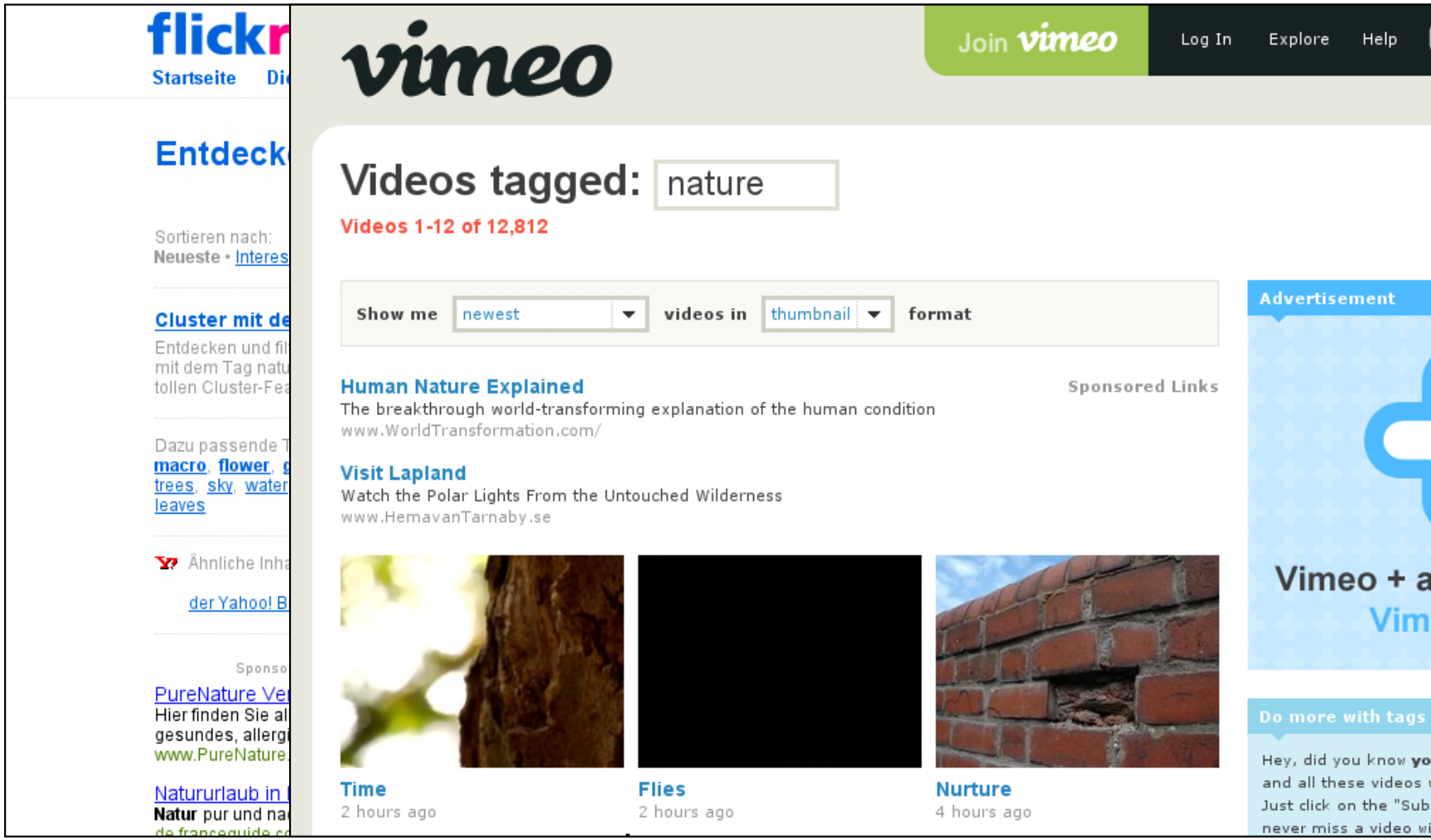
Agenda

6. Triadic Formal Concept Analysis
   - Motivation
   - Folksonomies
   - Triadic Formal Concepts
   - Concept-Tri-Lattice
   - Visualization of Tri-Lattices
   - Iceberg Tri-Lattices
   - Computing Tri-Concepts
   - Evaluation
   - Neighborhoods

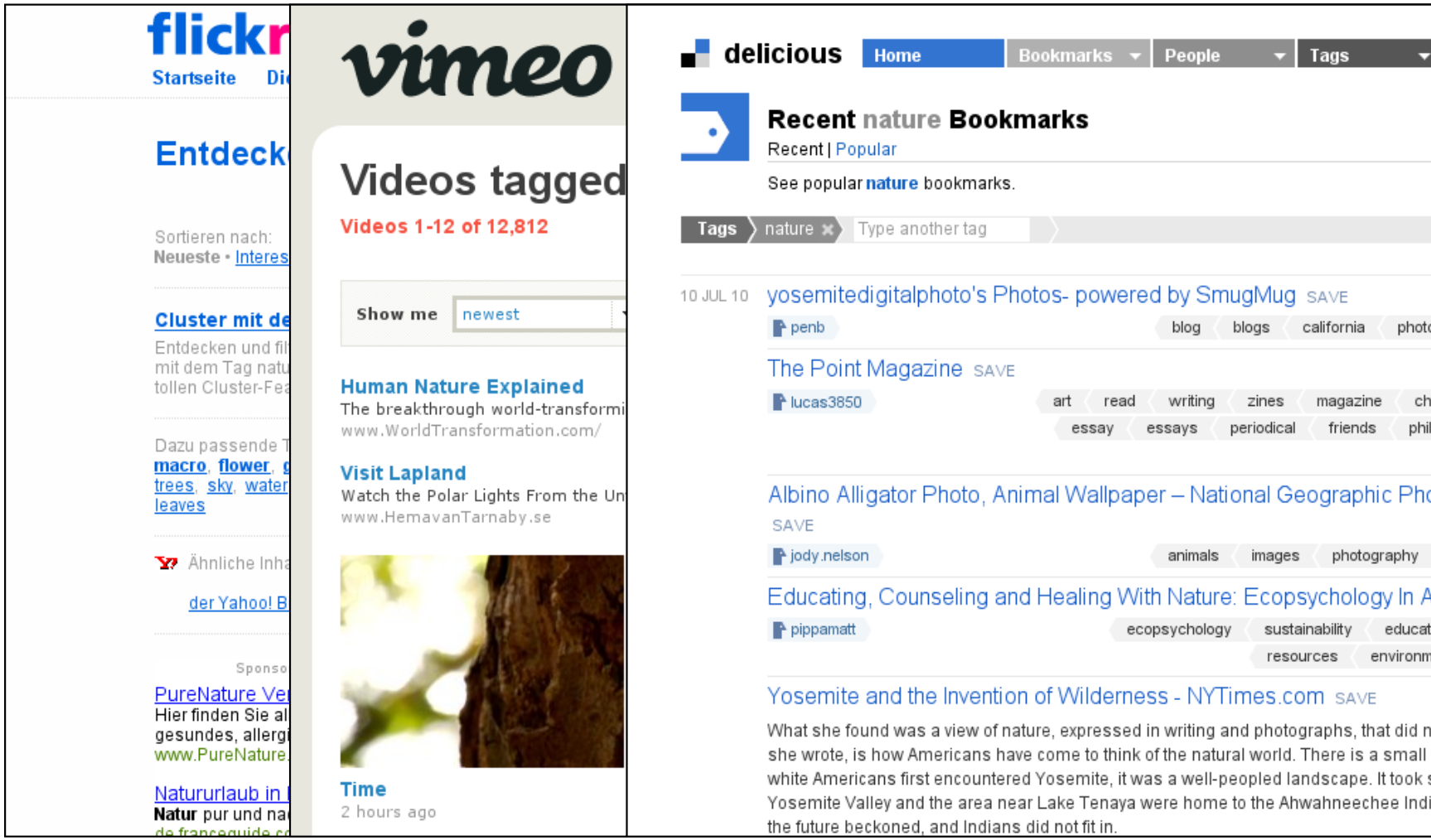
Motivation: Collaborative Tagging Systems



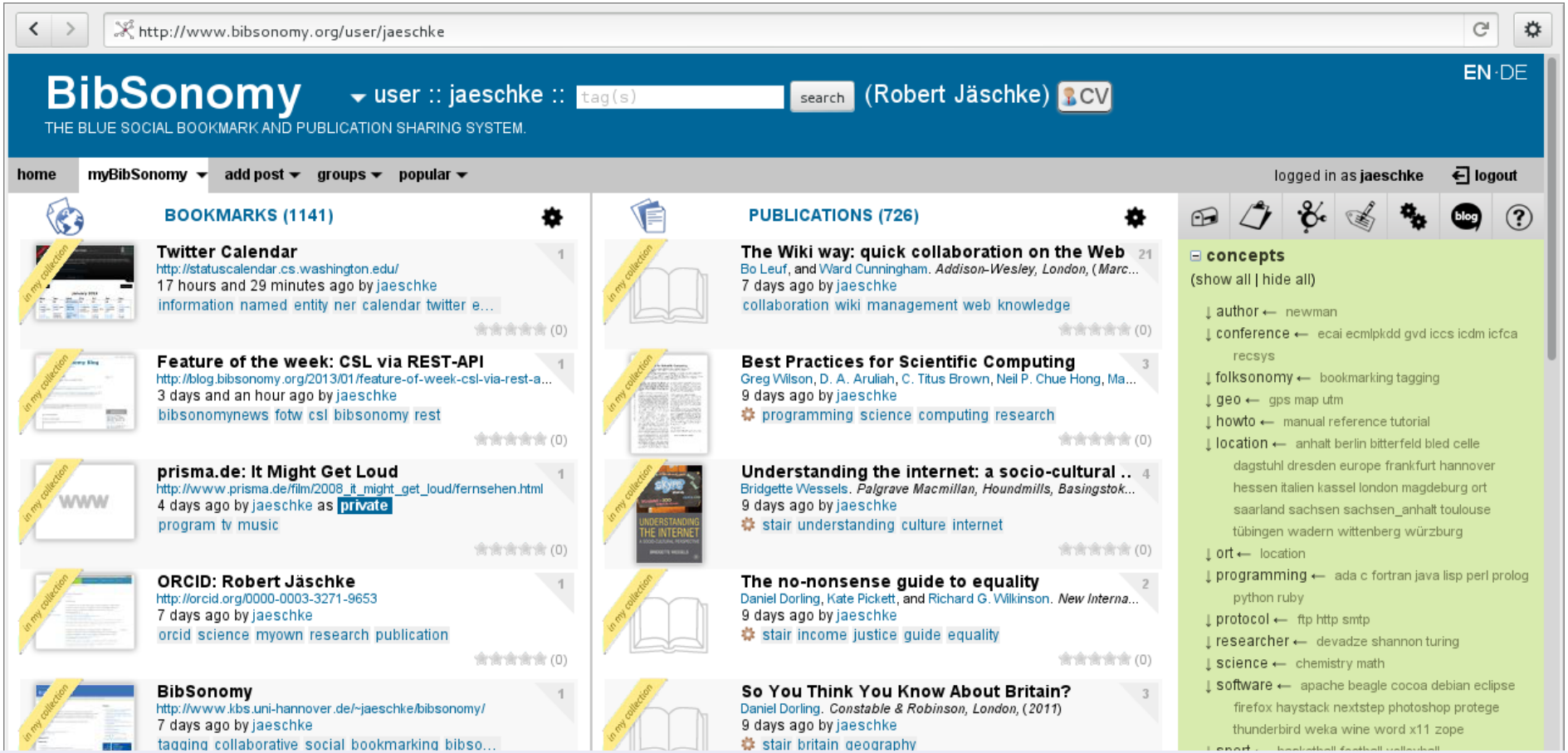
Motivation: Collaborative Tagging Systems



Motivation: Collaborative Tagging Systems



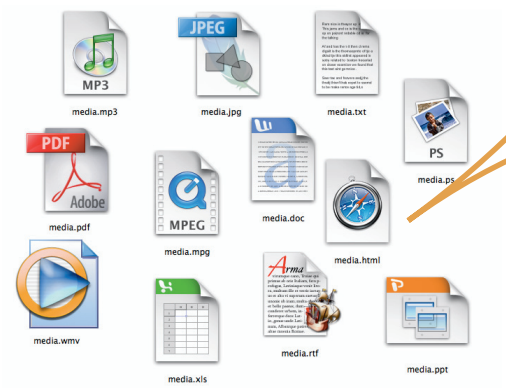
Motivation: Collaborative Tagging Systems



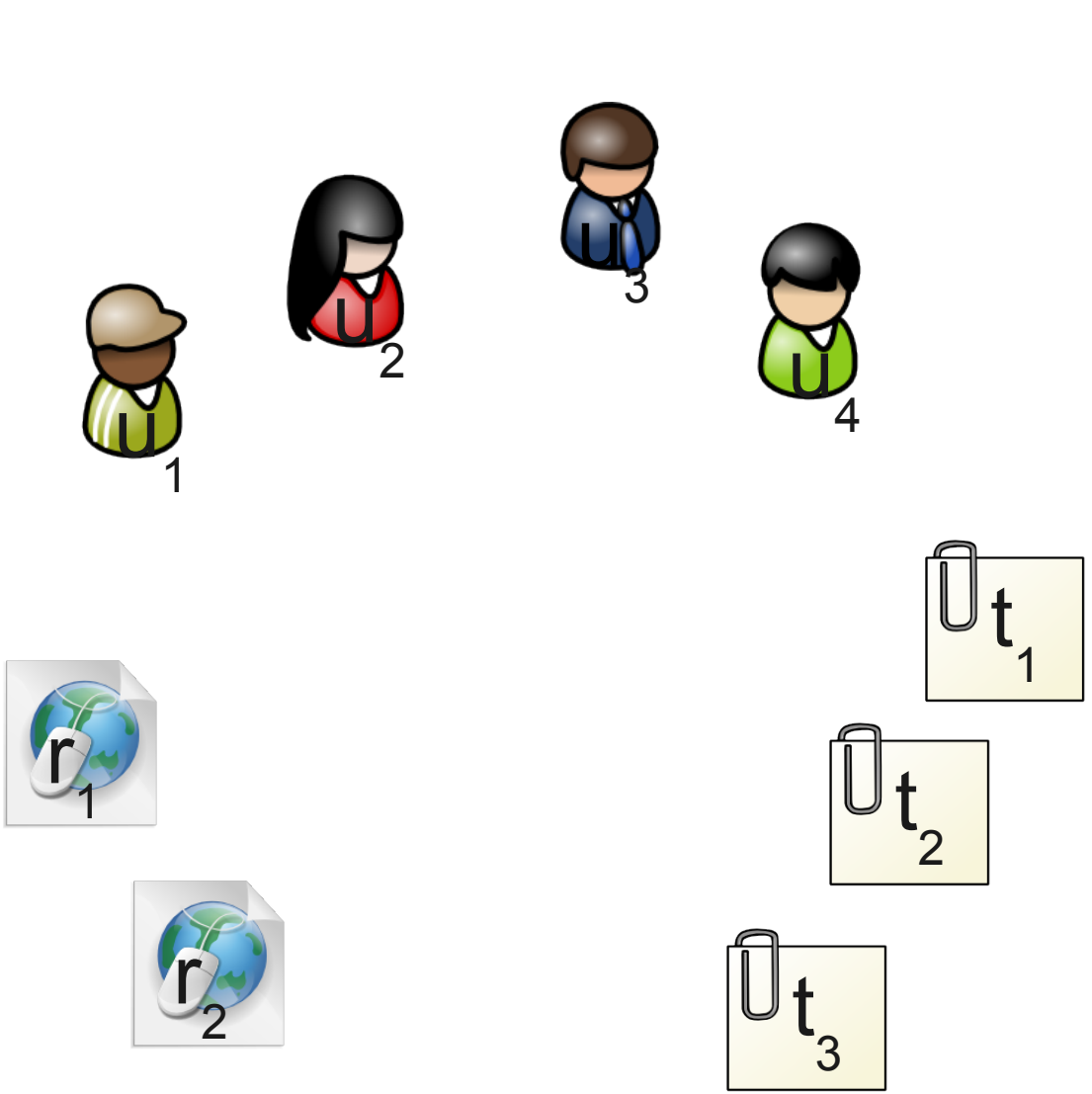
- manage your web bookmarks and publication references
- open for the public since beginning of 2006, $>$ 5 000 active users
- developed and operated at L3S Research Center

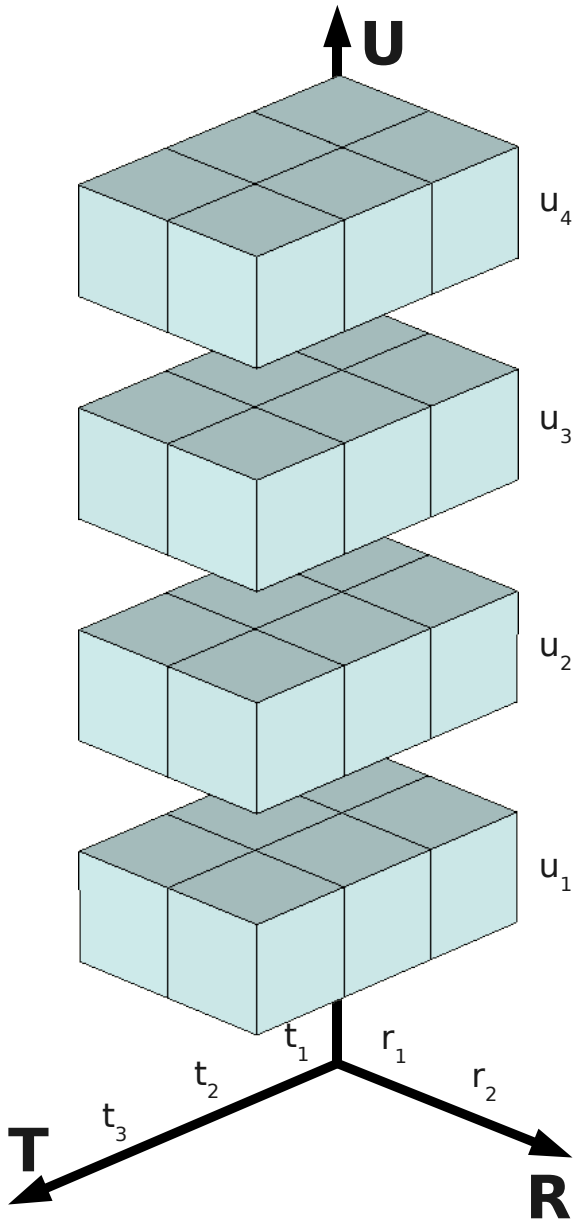
Folksonomies

- data structure of collaborative tagging systems
- connects users, tags, and resources
- conceptual structure created by the people

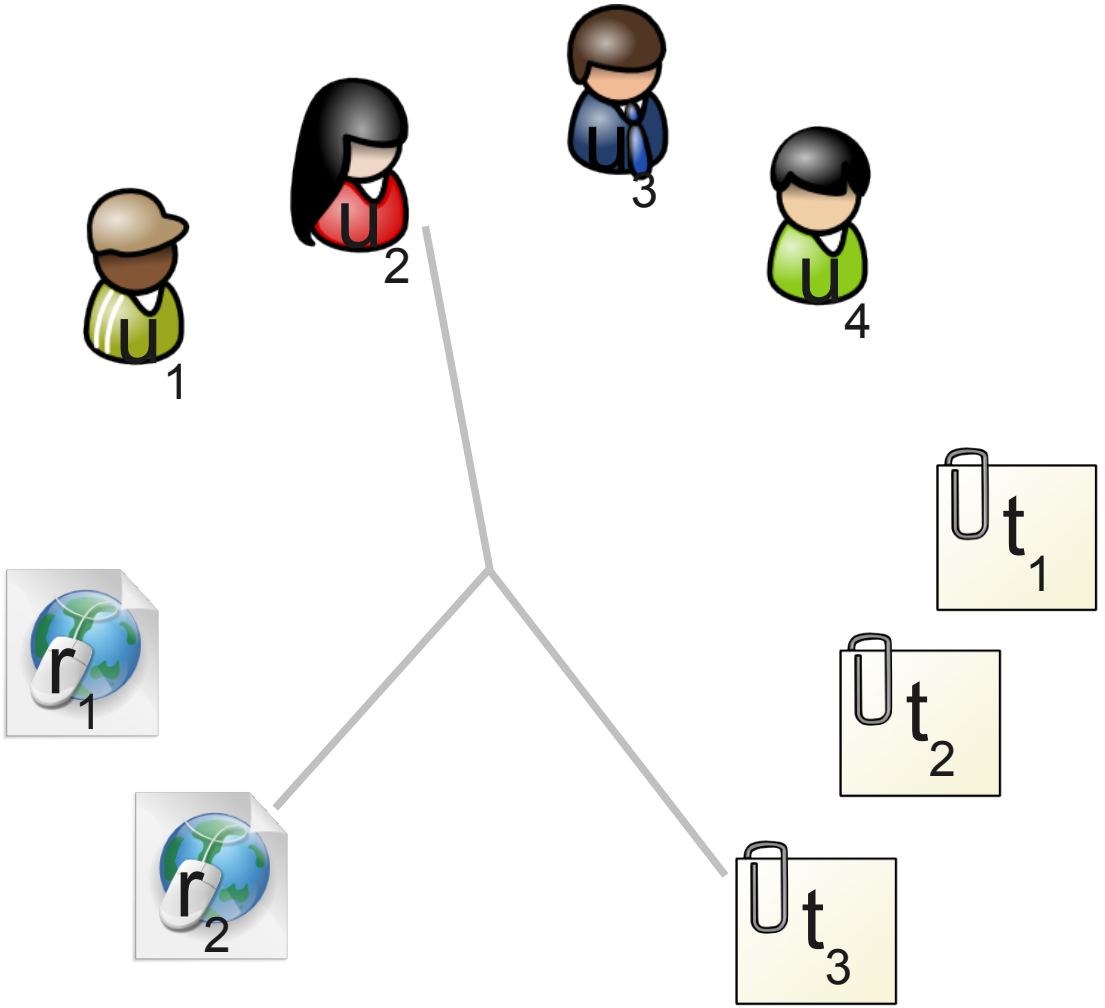


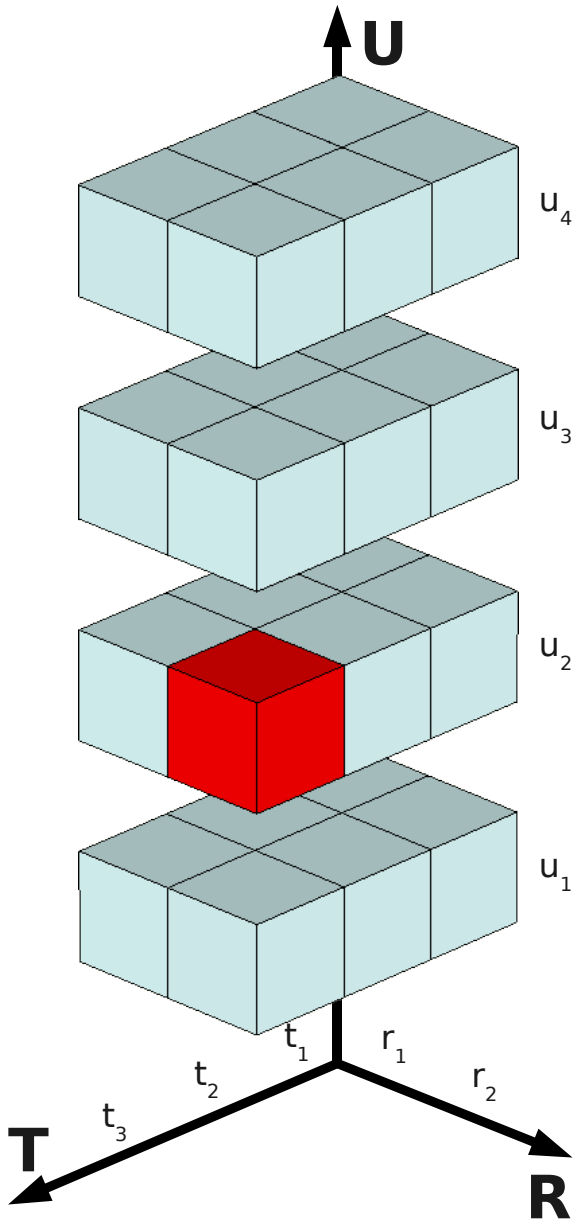
Folksonomies: Hypergraph, Tensor



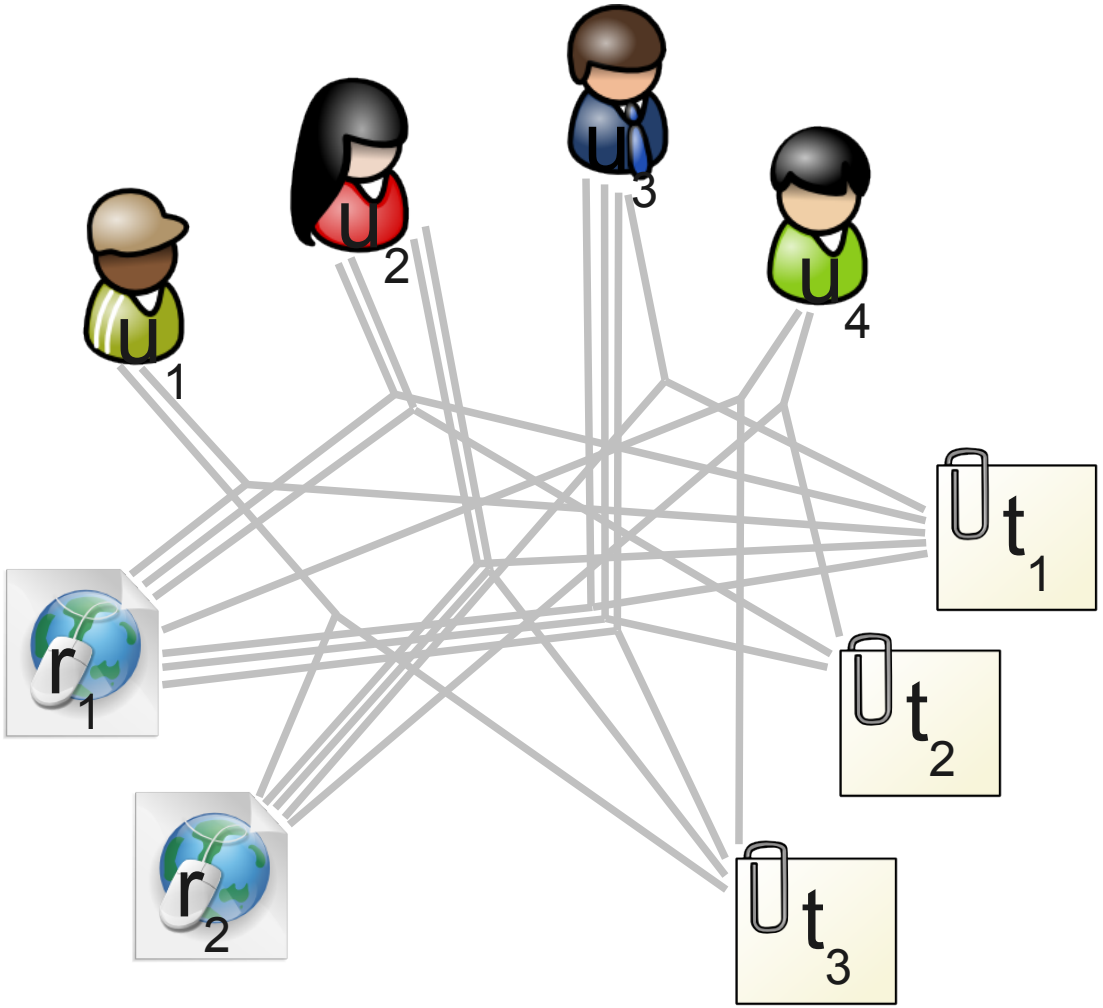


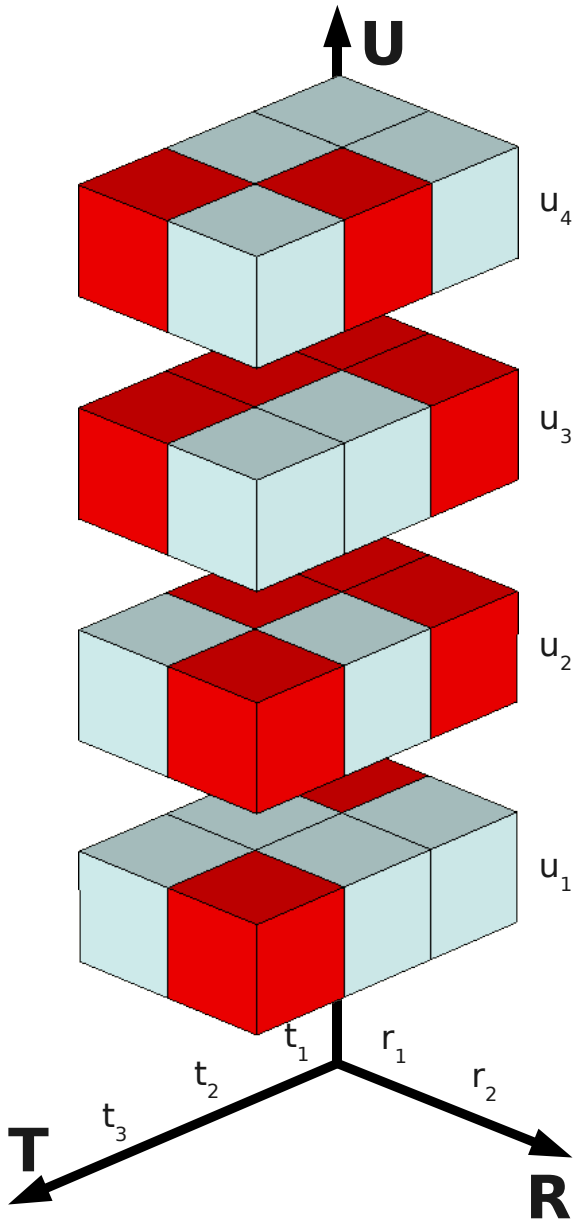
Folksonomies: Hypergraph, Tensor





Folksonomies: Hypergraph, Tensor





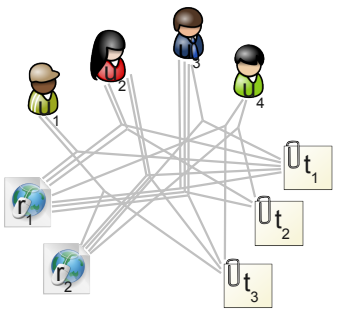
Folksonomies

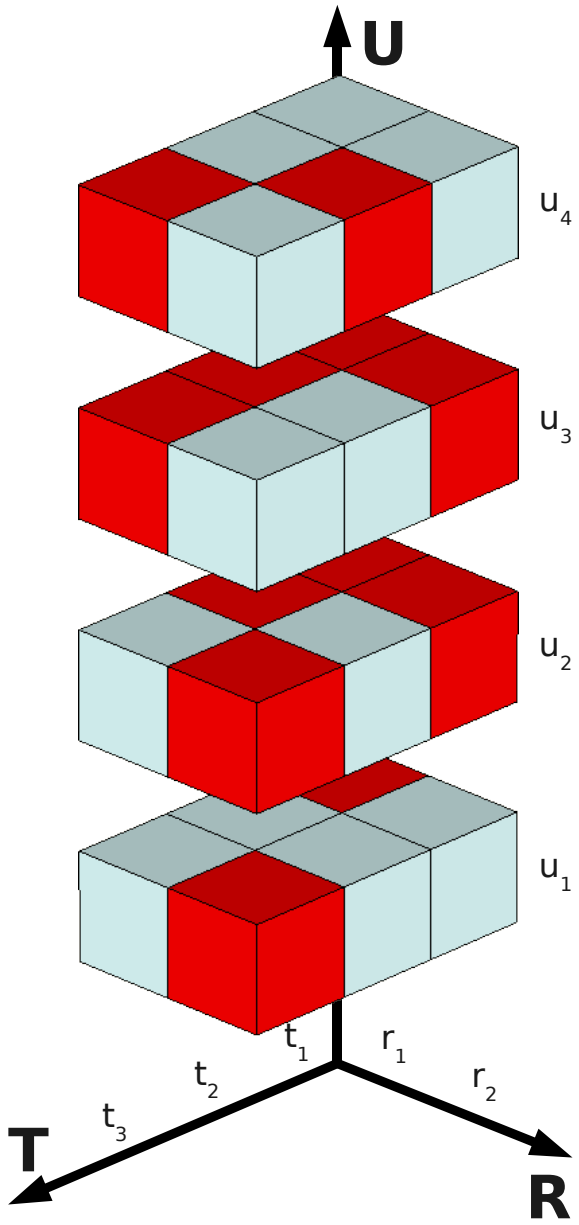
Definition (Folksonomy)

$\mathbb{F} := (U, T, R, Y)$ with

- U, T, R finite sets of users, tags, and resources, resp.
- $Y \subseteq U \times T \times R$ ternary relation

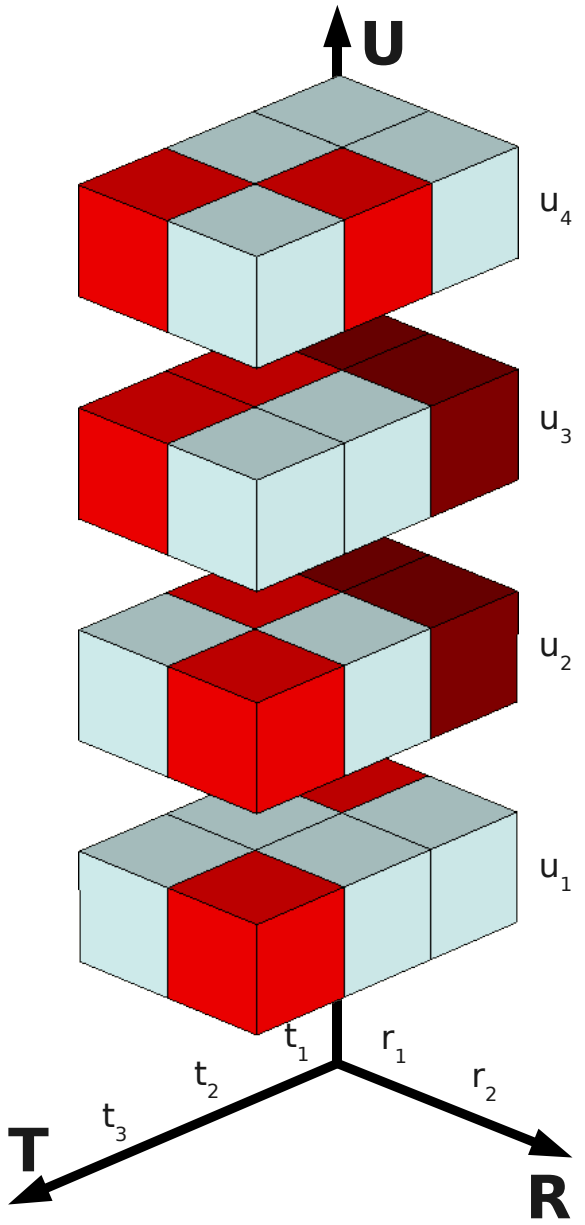
- tripartite hypergraph
- boolean 3-dimensional tensor
- triadic formal context





Folksonomies

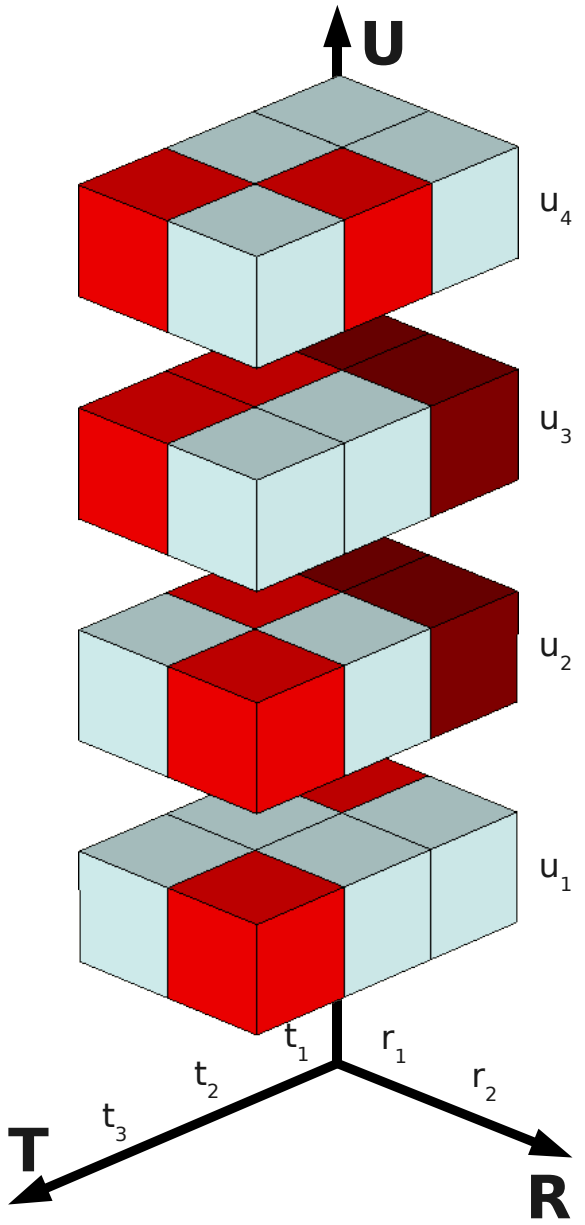
- conceptual clustering of folksonomies
  - find interessting concepts/clusters
  - support browsing, community detection, recommendations
  - get an overview into the structure of a folksonomy



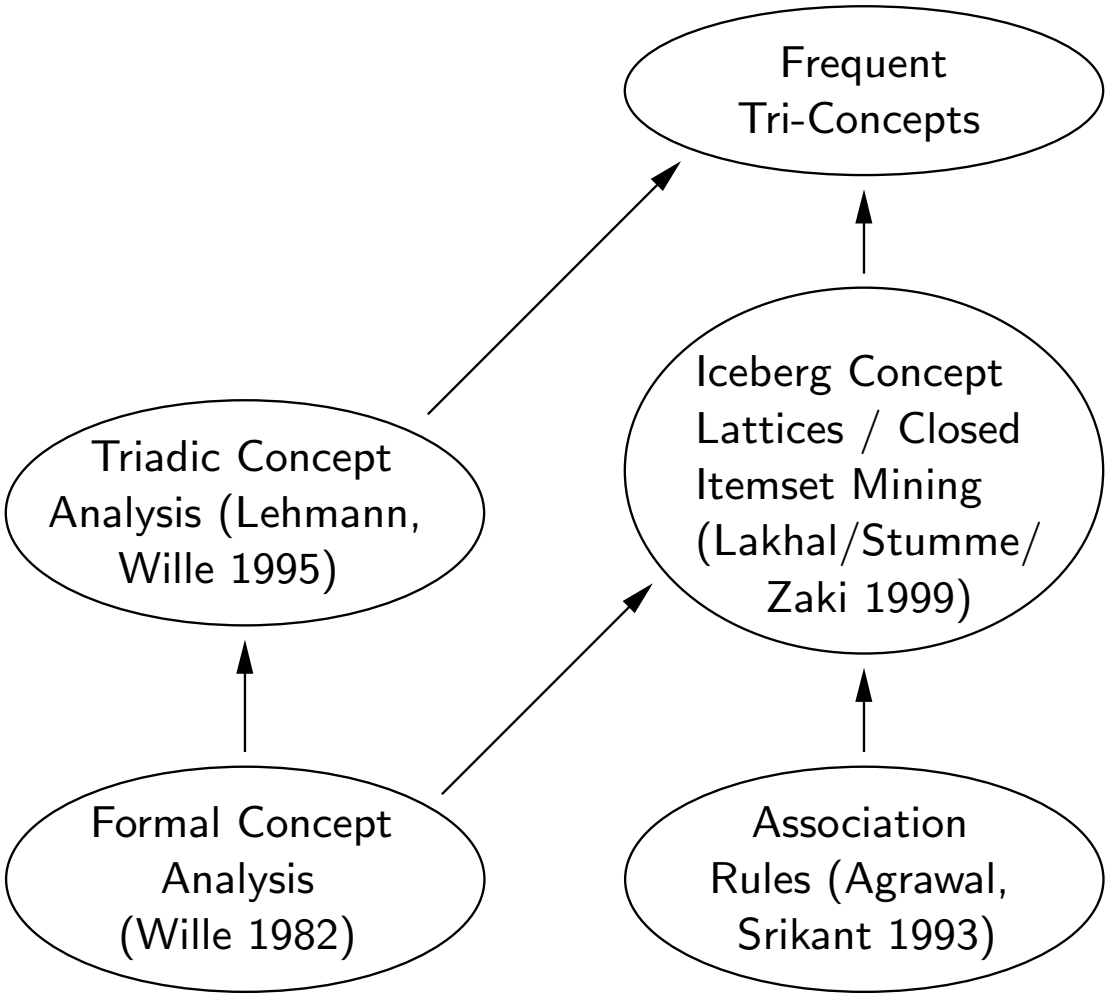
Folksonomies

- conceptual clustering of folksonomies
  - find interessting concepts/clusters
  - support browsing, community detection, recommendations
  - get an overview into the structure of a folksonomy
- *tri-concept* $(A, B, C) \subseteq U \times T \times R$: maximal cuboid in which every user from A has tagged every ressource from C with all tags from B

  → shared conceptualization

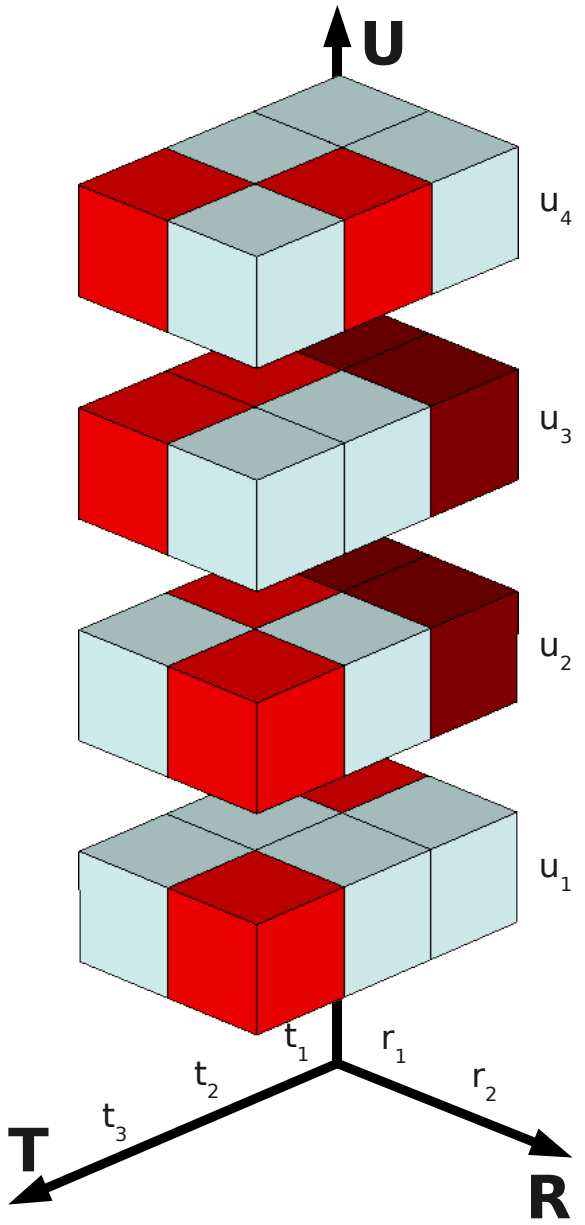


Folksonomies



Folksonomies

We regard $\mathbb{F} = (U, T, R, Y)$ as *triadic formal context*.
In general, the elements of U, T and R are then called *objects*, *attributes* and *conditions* and $(u, t, r) \in Y$ is read as "*object* u *has the attribute* t *under condition* r".

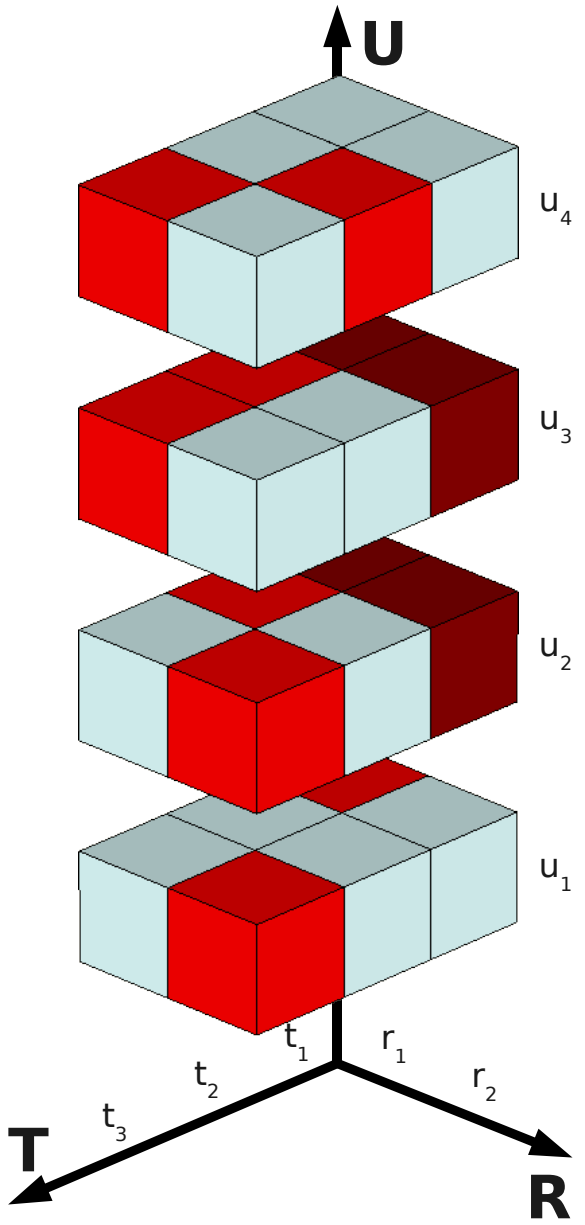


Triadic Formal Concepts

Definition (tri-concept)

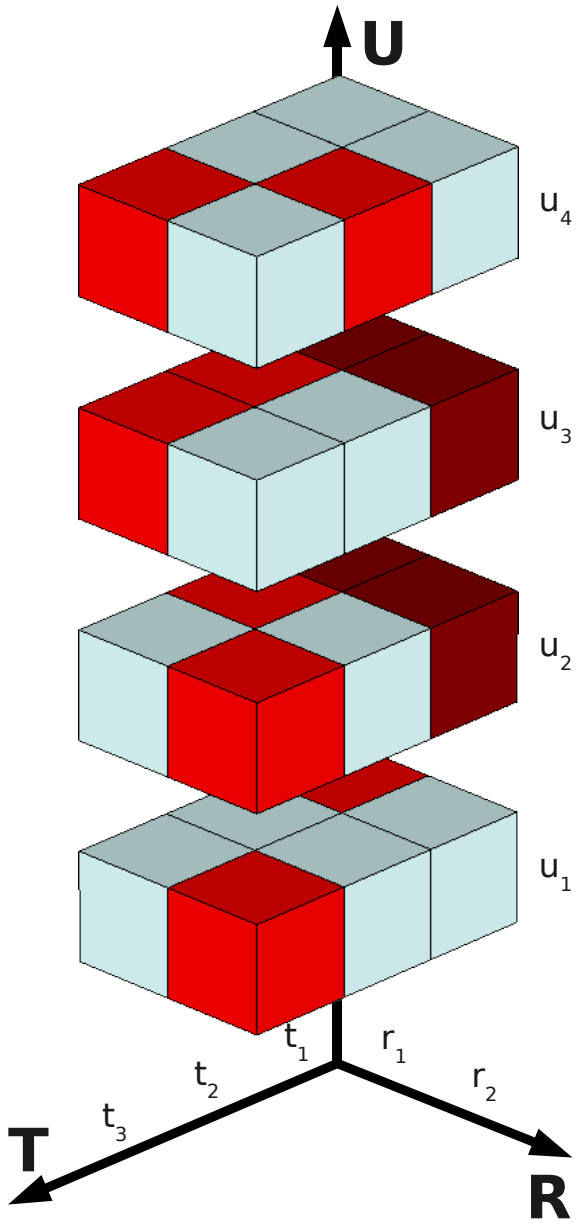
triple (A, B, C) with $A \subseteq U$, $B \subseteq T$, $C \subseteq R$ and $A \times B \times C \subseteq Y$, such that none of the three components can be enlarged without violating the condition $A \times B \times C \subseteq Y$.
We call A the *extent*, B the *intent* and C the *modus* of the formal tri-concept.

→ natural extension of formal concepts



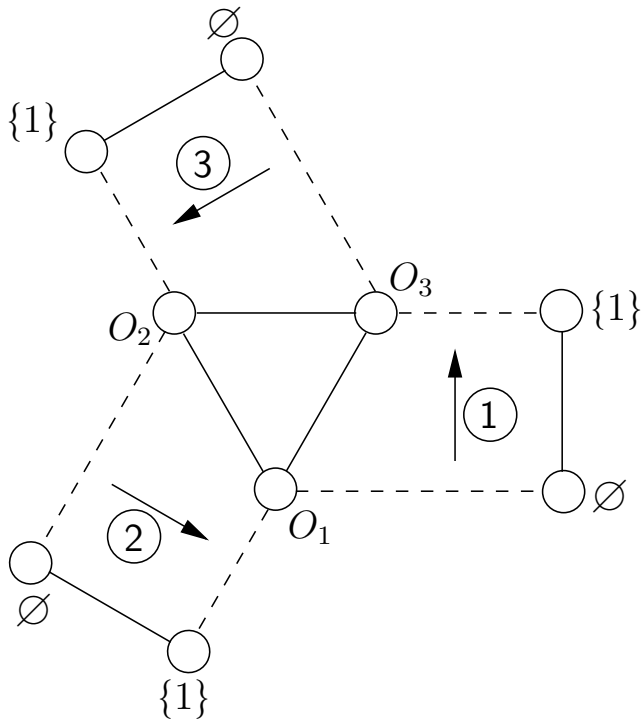
Concept-Tri-Lattice

- three quasi orders $\lesssim_1$, $\lesssim_2$, $\lesssim_3$:
  $(A_1, A_2, A_3) \lesssim_i (B_1, B_2, B_3)$
  $:\Leftrightarrow\ A_i \subseteq B_i$, for $i = 1, 2, 3$.
- *not antisymmetric*, i. e. from
  $(A_1, A_2, A_3) \lesssim_i (B_1, B_2, B_3)$ and
  $(B_1, B_2, B_3) \lesssim_i (A_1, A_2, A_3)$ does not
  follow $(A_1, A_2, A_3) = (B_1, B_2, B_3)$
- *concept tri-lattice* $\underline{\mathfrak{B}}(\mathbb{K})$ of the triadic context $\mathbb{K}$
- not a real (mathematical) lattice!

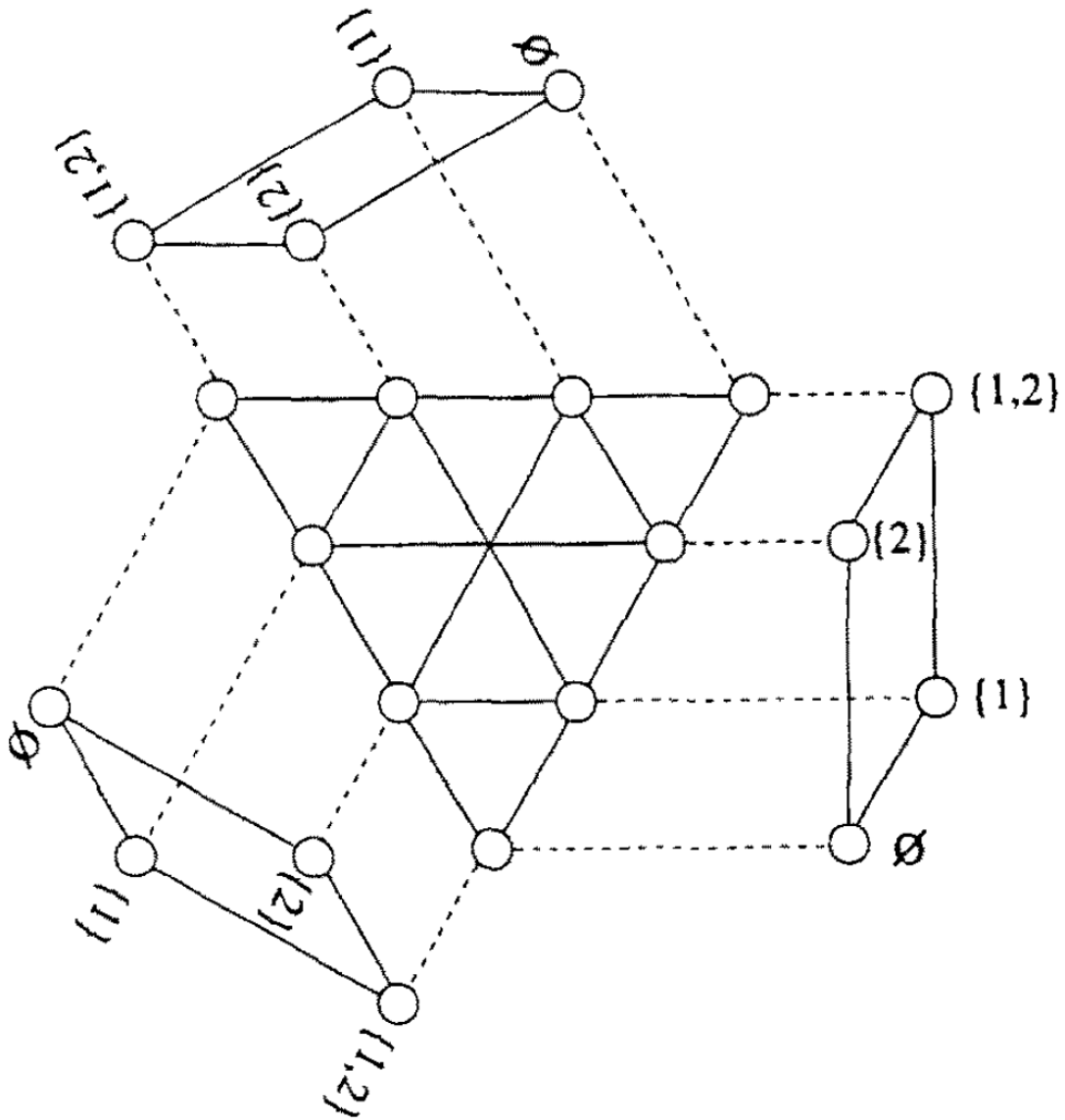


Visualization of Tri-Lattices

- Since it is not really a lattice, we can not draw a lattice diagram
- Alternative:
  - every quasi-order is written along the edge of a virtual triangle
  - the tri-concepts are drawn into the triangle
- example to the right: smallest non-trivial tri-lattice
  $\mathbb{B}_3 = \underline{\mathfrak{B}}(\{1\}, \{1\}, \{1\}, \varnothing)$
- visualization not always possible
  - satisfied tetrahedron condition
  - violated Thomson condition



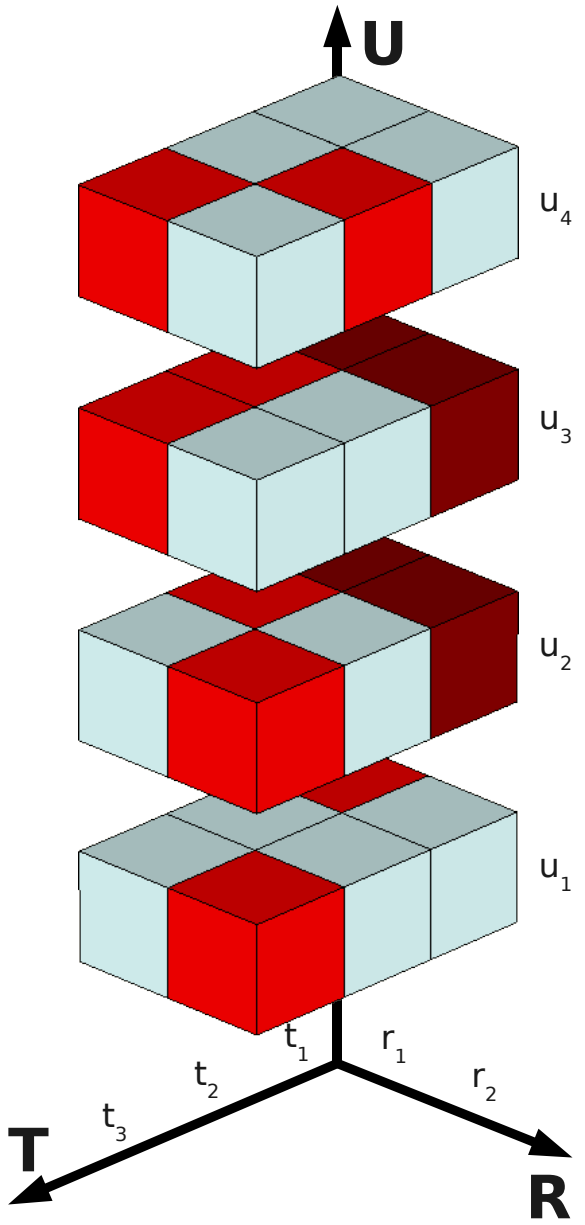
Visualization of Tri-Lattices



Iceberg Tri-Lattices

- Given support constraints τ_u, τ_t, τ_r : tri-concept (A, B, C) *frequent* $:\Leftrightarrow$ $|A| \geqslant \tau_u$, $|B| \geqslant \tau_t$, and $|C| \geqslant \tau_r$

  $\rightarrow$ *iceberg tri-lattice*

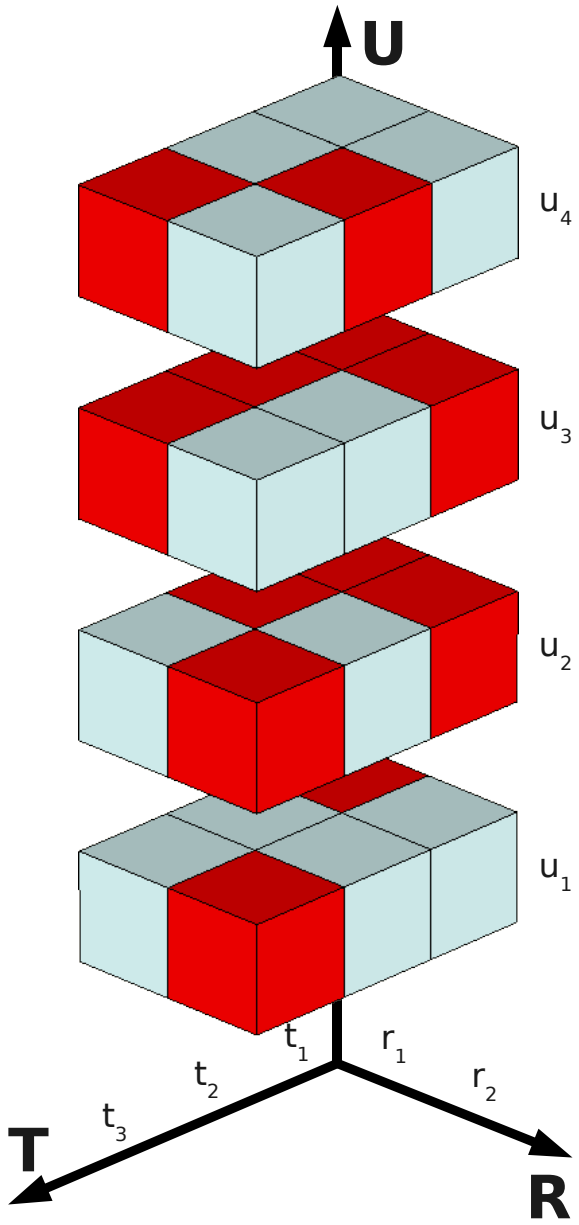


Computing Tri-Concepts

- Given
  - sets U, T, R
  - ternary relation $Y \subseteq U \times T \times R$
  - support constraints τ_u, τ_t, τ_r
- Find (A, B, C) with
  - $A \subseteq U$, $B \subseteq T$, $C \subseteq R$
  - $|A| \geqslant \tau_u$, $|B| \geqslant \tau_t$, $|C| \geqslant \tau_r$
  - $A \times B \times C \subseteq Y$
  - such that none of the sets A, B or C can be enlarged without violating the former condition

Computing Tri-Concepts

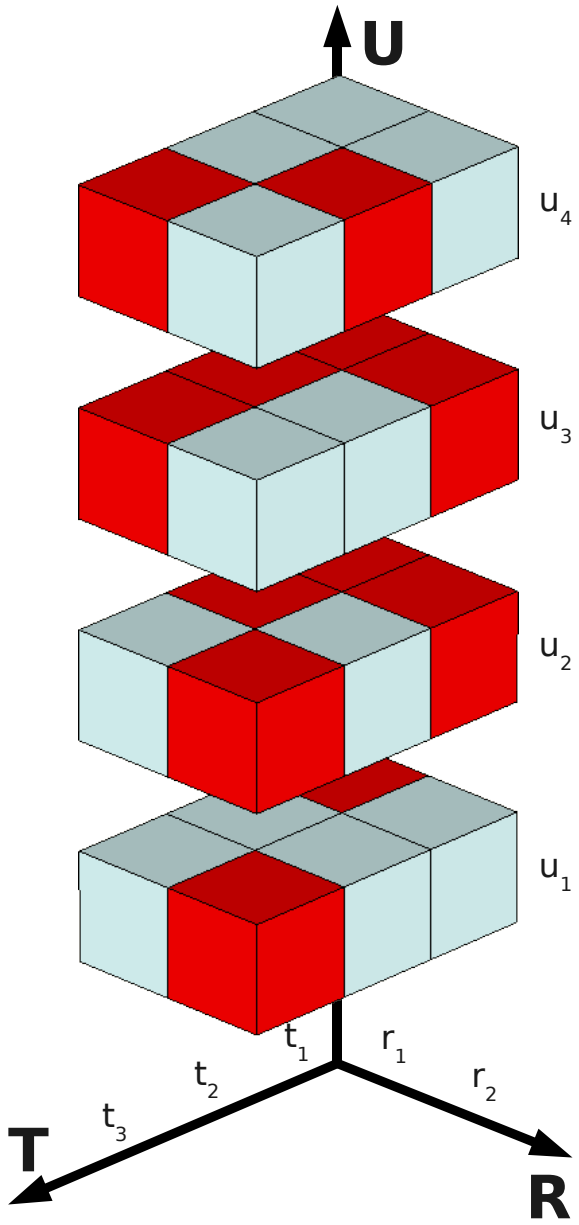
computes the iceberg tri-lattice of a triadic formal context



Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

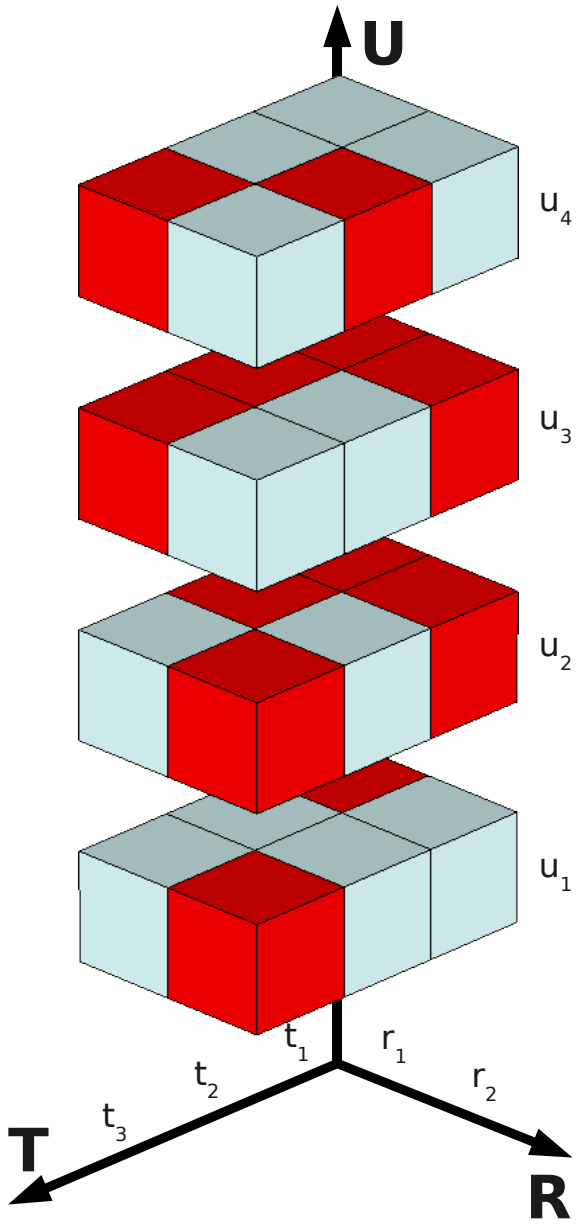


Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u, (t, r)) \mid (u, t, r) \in Y\}$

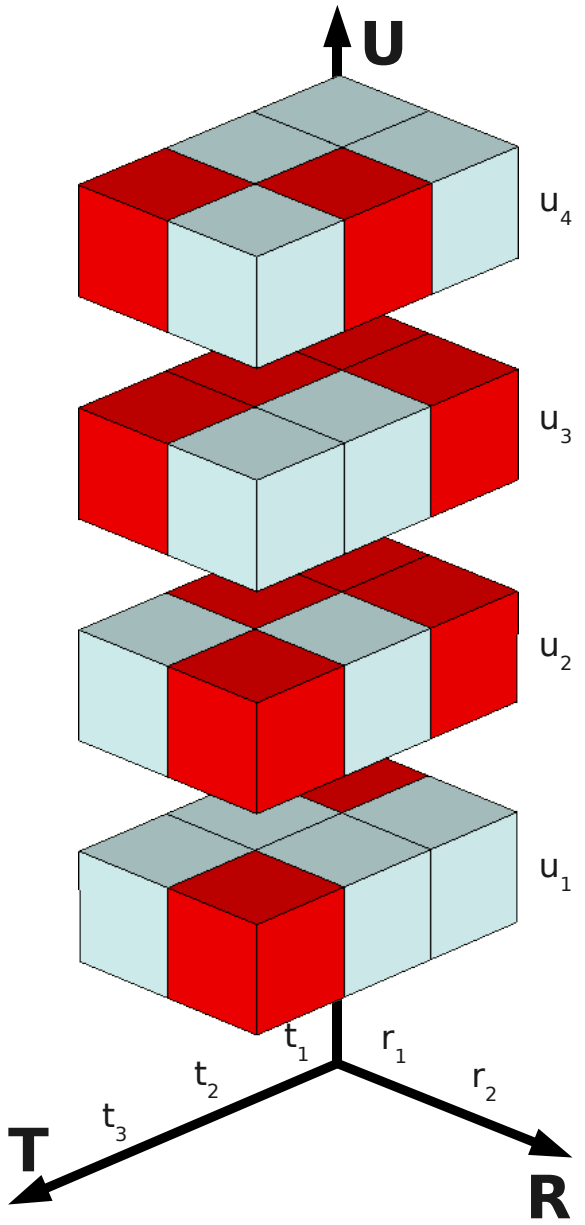


Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$



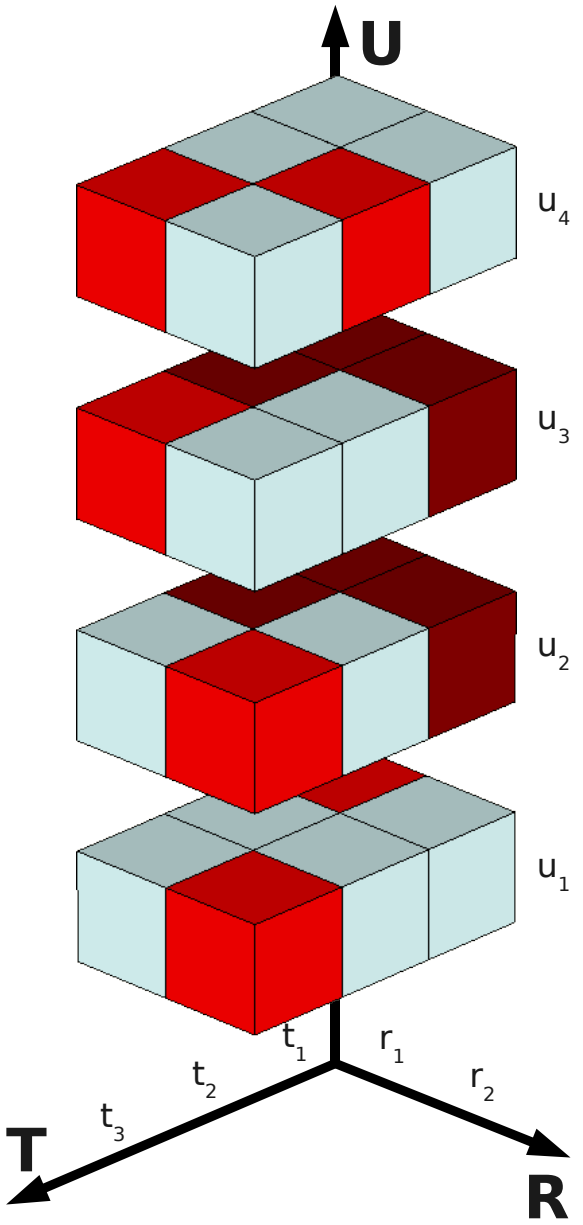
Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$

In the example:
$(A, I) =$
$\big(\{u_2, u_3\},\ \{(t_1, r_1), (t_1, r_2), (t_2, r_1)\}\big)$



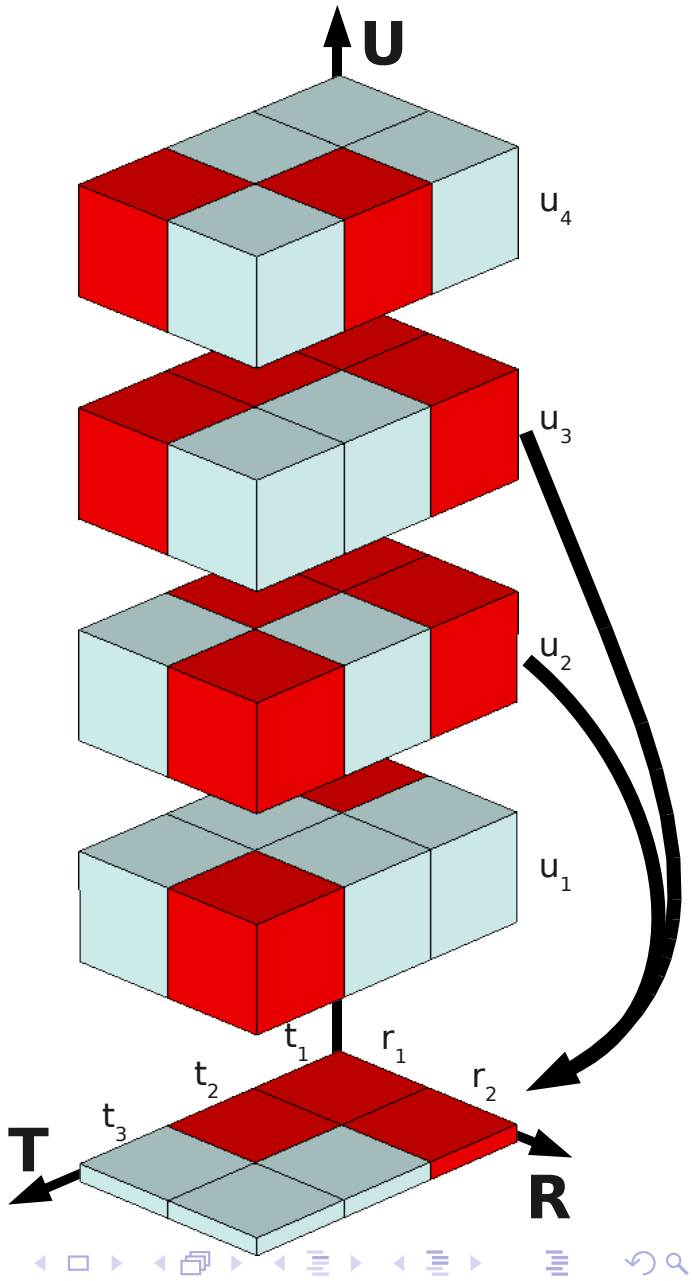
Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$
  - Loop: Find (frequent) concepts $(\mathbf{B}, \mathbf{C})$ in (T, R, I)

In the example:
$(T,R,I) = (T,R,\{(t_1,r_1),(t_1,r_2),(t_2,r_1)\})$



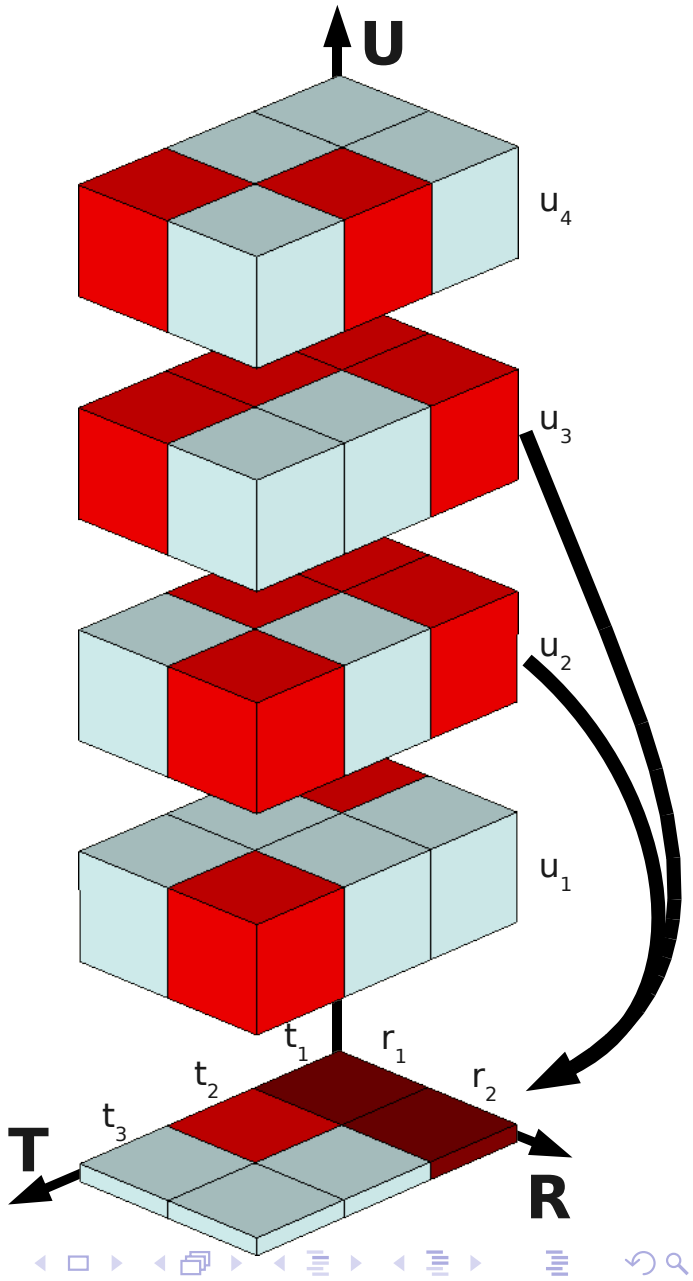
Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$
  - Loop: Find (frequent) concepts $(\mathbf{B}, \mathbf{C})$ in (T, R, I)

In the example:
$(B,C) = (\{t_1\}, \{r_1, r_2\})$



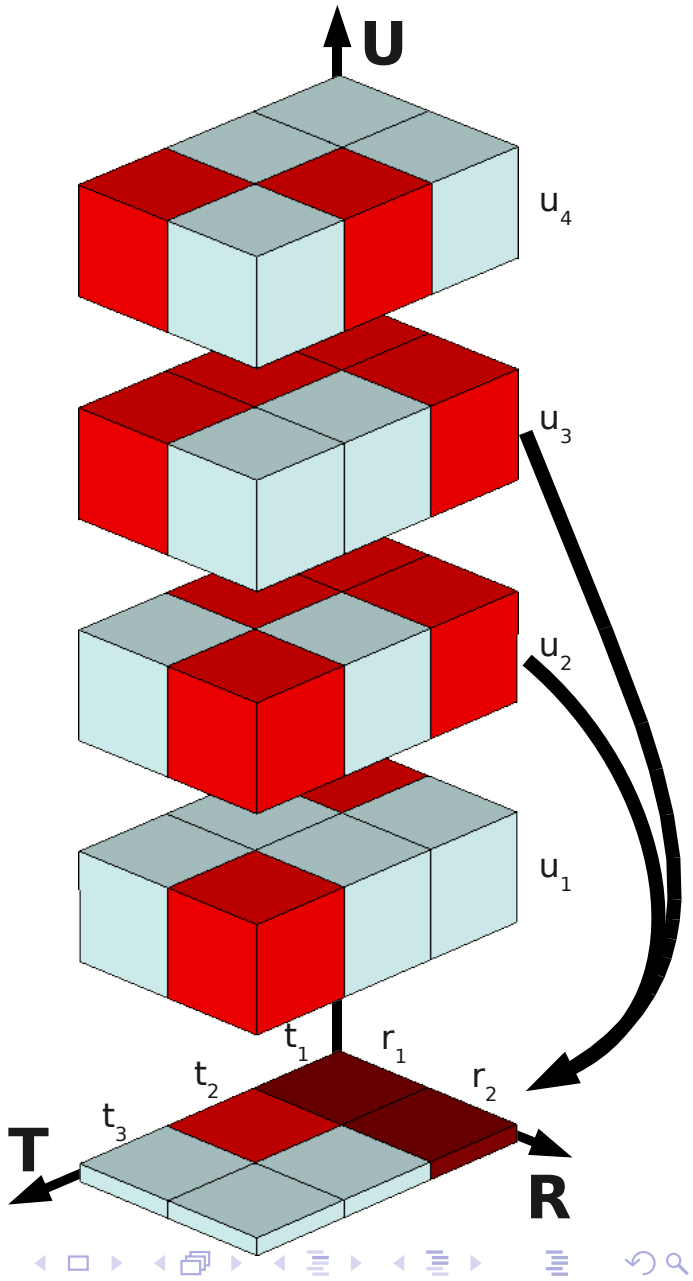
Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$
  - Loop: Find (frequent) concepts $(\mathbf{B}, \mathbf{C})$ in (T, R, I)
    - If $\mathbf{A} = (\mathbf{B} \times \mathbf{C})^{\tilde{Y}}$, then output $(\mathbf{A}, \mathbf{B}, \mathbf{C})$

In the example:
$(B \times C)^{\tilde{Y}} = (\{t_1\} \times \{r_1, r_2\})^{\tilde{Y}}$



Computing Tri-Concepts

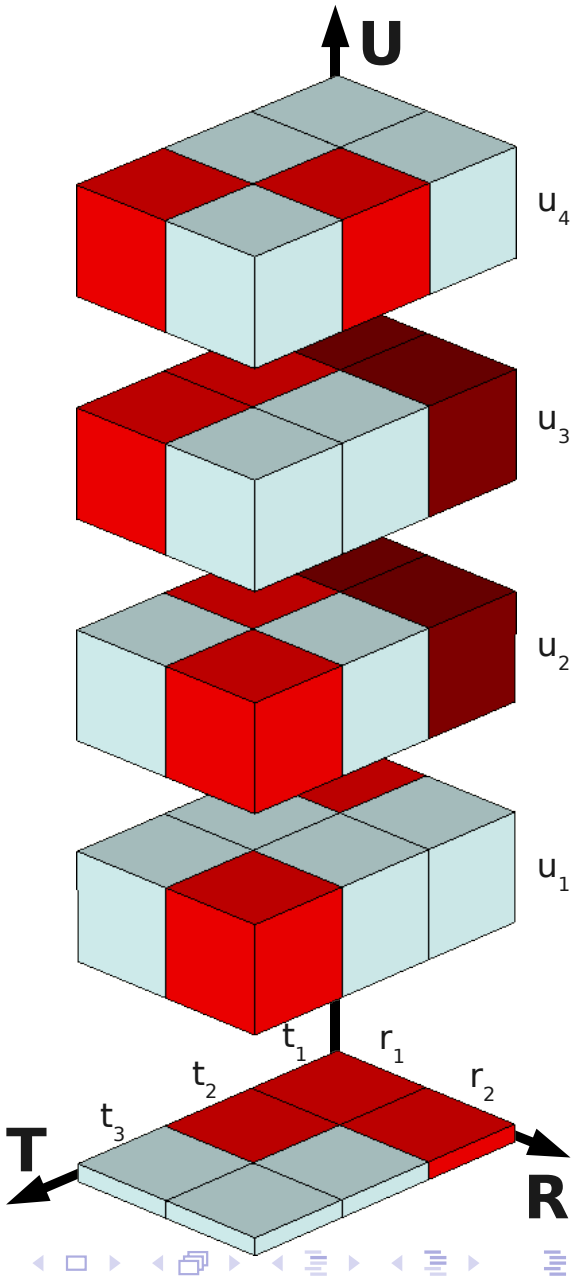
computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$
  - Loop: Find (frequent) concepts $(\mathbf{B}, \mathbf{C})$ in (T, R, I)
    - If $\mathbf{A} = (\mathbf{B} \times \mathbf{C})^{\tilde{Y}}$, then output $(\mathbf{A}, \mathbf{B}, \mathbf{C})$

In the example:

$$(B \times C)^{\tilde{Y}} = (\{t_1\} \times \{r_1, r_2\})^{\tilde{Y}} = \{u_2, u_3\} = A$$



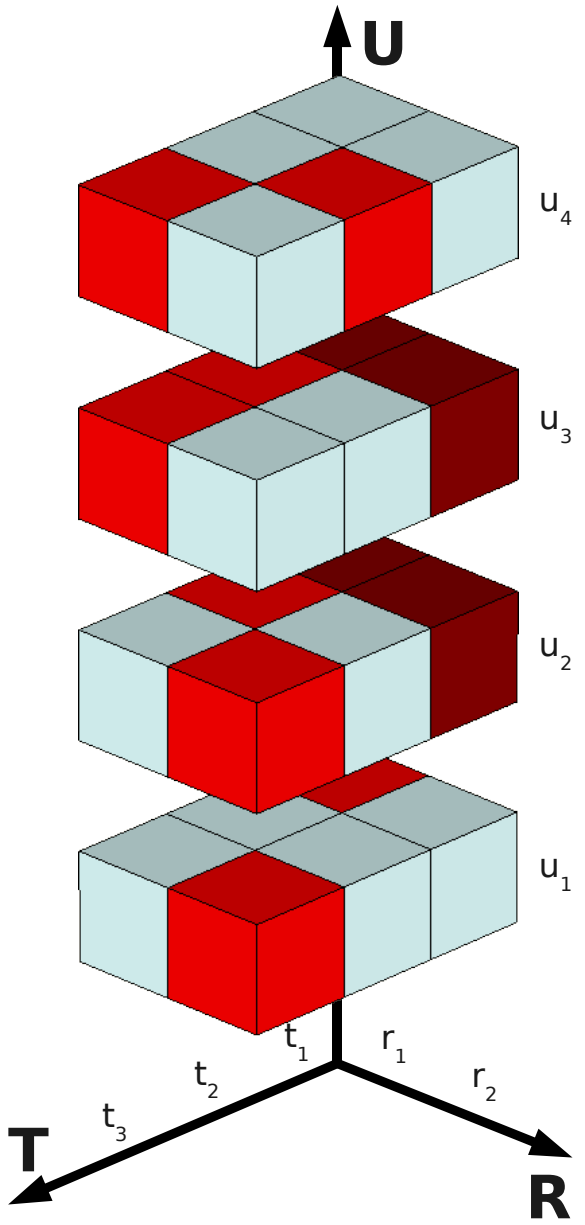
Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

- Let $\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
- Loop: Find (frequent) concepts $(\mathbf{A}, I)$ in $(U, T \times R, \tilde{Y})$
  - Loop: Find (frequent) concepts $(\mathbf{B}, \mathbf{C})$ in (T, R, I)
    - If $\mathbf{A} = (\mathbf{B} \times \mathbf{C})^{\tilde{Y}}$, then output $(\mathbf{A}, \mathbf{B}, \mathbf{C})$

In the example:
$(A, B, C) = (\{u_2, u_3\}, \{t_1\}, \{r_1, r_2\})$



Computing Tri-Concepts

Require: $U, T, R, Y, \tau_u, \tau_t, \tau_r$

$\tilde{Y} := \{(u,(t,r)) \mid (u,t,r) \in Y\}$
$(A, I) :=$ *FirstFrequentConcept*$((U, T \times R, \tilde{Y}), \tau_u)$
repeat
  if $|I| \geqslant \tau_t \cdot \tau_r$ **then**
    $(B, C) :=$ *FirstFrequentConcept*$((T, R, I), \tau_t)$
    repeat
      if $|C| \geqslant \tau_r$ **then**
        if $A = (B \times C)^{\tilde{Y}}$ **then**
          print A,B,C
        end if
      end if
    until not *NextFrequentConcept*$((B, C), (T, R, I), \tau_t)$
  end if
until not *NextFrequentConcept*$((A, I), (U, T \times R, \tilde{Y}), \tau_u)$

Computing Tri-Concepts

The *FirstFrequentConcept* method:

Require: $(G, M, I), \tau$

1: $A := \varnothing^I$
2: $B := A^I$
3: **if** $|A| < \tau$ **then**
4: *NextFrequentConcept*$((A, B), (G, M, I), \tau)$
5: **end if**
6: **return** (A, B)

Computing Tri-Concepts

the *NextFrequentConcept* method:

Require: $(A, B), (G, M, I), \tau$

1: i := *max*(M)
2: **while** *defined*(i) **do**
3: $A := (B \bullet i)^I$
4: **if** $|A| \geqslant \tau$ **then**
5: $D := A^I$
6: **if** $B <_i D$ **then**
7: $B := D$
8: **return** **true**
9: **end if**
10: **end if**
11: $i :=$ *max*$(M \backslash B \cap \{1, \ldots, i-1\})$
12: **end while**
13: **return** **false**

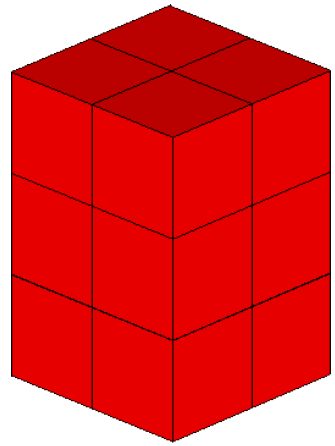
Evaluation

BibSonomy Dataset:

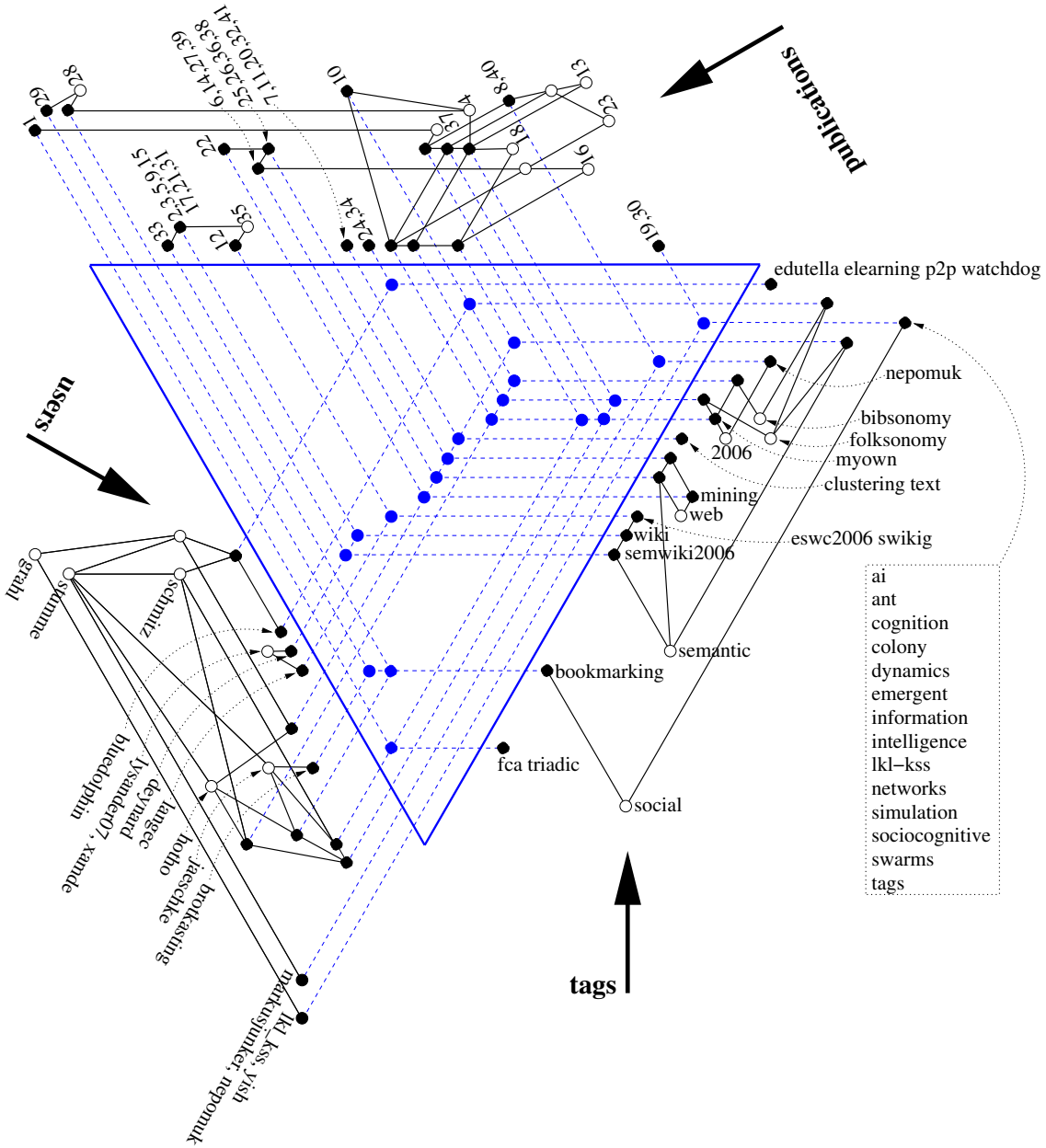
- all publication records until November 23rd, 2006
- removed: DBLP, posts with the tag "imported"
- $|U| = 262$, $|T| = 5\,954$, $|R| = 11\,101$, $|Y| = 44\,944$

Result:

- 13 992 tri-concepts (75 minutes on a 2 GHz PC)
- with support constraints $\tau_u = 3$, $\tau_t = 2$, $\tau_r = 2$:
  - 21 tri-concepts
  - contain 41 publications, 15 users and 36 tags

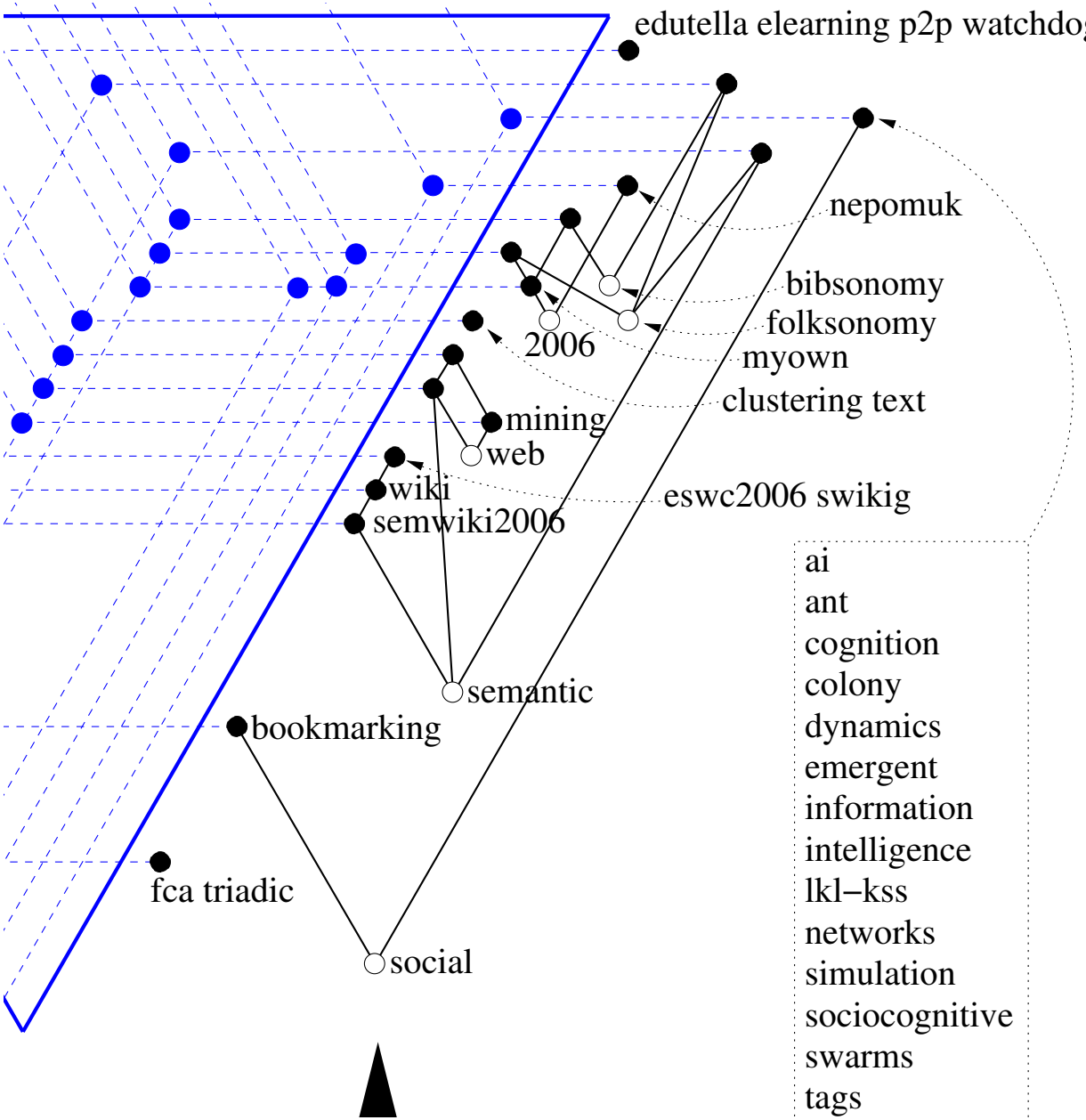


Evaluation



visualisation of the iceberg tri-lattice for $\tau_u = 3$, $\tau_t = 2$, $\tau_r = 2$

Evaluation



two topical groups:

- semantic
- social

semantic further divided:

- wiki
- web
- folksonomy

Neighborhoods

The visualization of tri-lattices is . . .

- at the moment manual work,
- time-intensive and pretty complicated,
- or even impossible (cf. *tetrahedron condition* and *Thomson condition*).

Thus: easier visualization option desireable

Neighborhoods

Idea:

- We regard tri-concepts as nodes in a graph.
- We connect two tri-concepts with an edge, when they contain the same tags, users, or resources.

More formally:

- Two tri-concepts (A_1, A_2, A_3) and (B_1, B_2, B_3) are *neighbors*, if for an $i \in \{1, 2, 3\}$ it holds $A_i = B_i$.
- neighbor relation $\sim \subseteq (\underline{\mathfrak{B}}(\mathbb{F}) \times \underline{\mathfrak{B}}(\mathbb{F}))$
- The *neighborhood graph* then is $(\underline{\mathfrak{B}}(\mathbb{F}), \sim)$.

Neighborhoods

neighborhood graph for the tri-concept
$(\{jaeschke, schmitz, stumme\}, \{fca, triadic\}, \{1, 37\})$