Formal Concept Analysis

III Knowledge Discovery

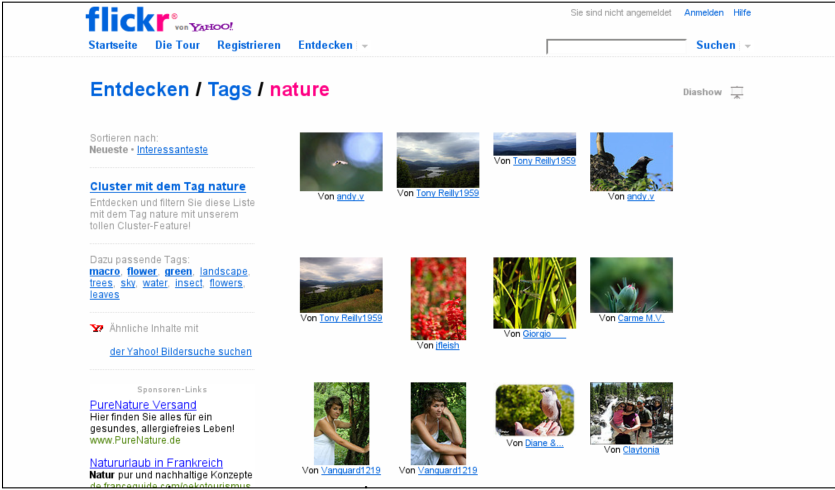
Sebastian Rudolph

Computational Logic Group
Technische Universität Dresden

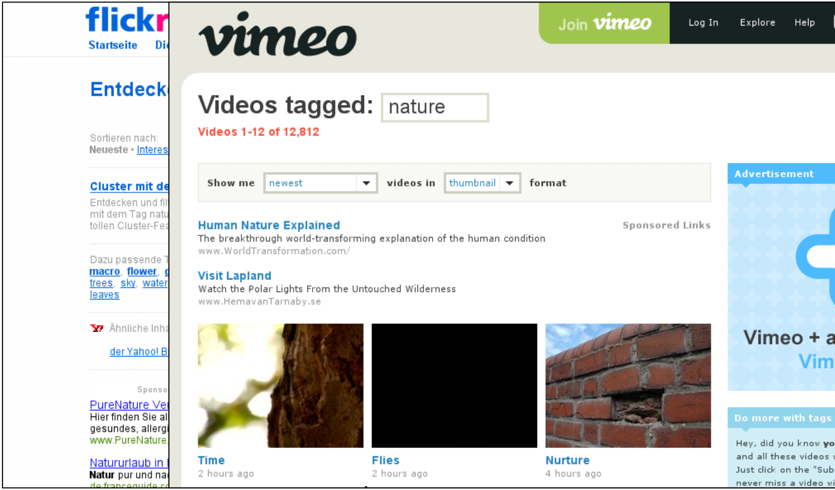
Agenda

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

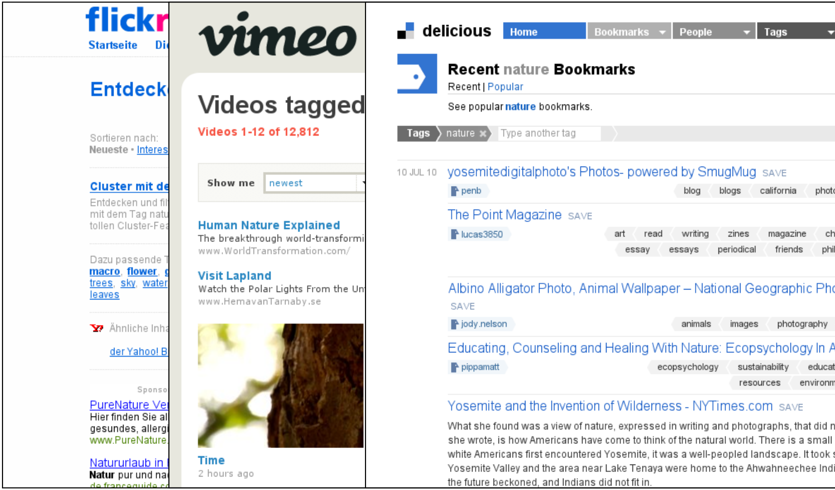
Motivation: Collaborative Tagging Systems



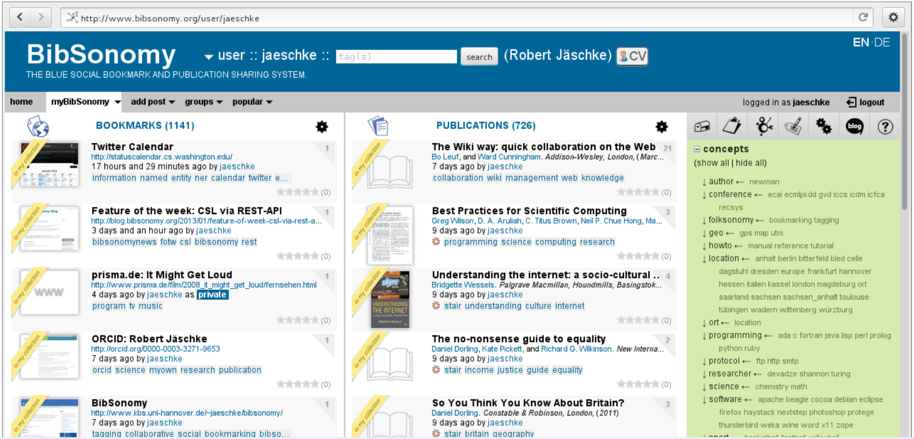
Motivation: Collaborative Tagging Systems



Motivation: Collaborative Tagging Systems



Motivation: Collaborative Tagging Systems



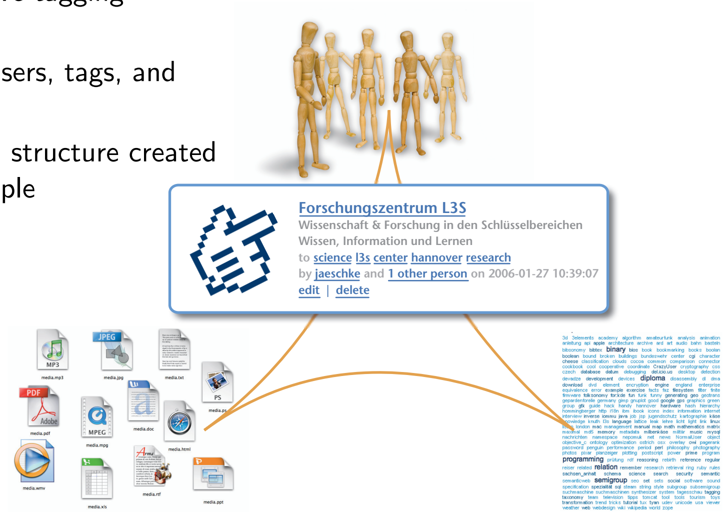
- manage your web bookmarks and publication references
- open for the public since beginning of 2006, > 5 000 active users
- developed and operated at L3S Research Center

Agenda

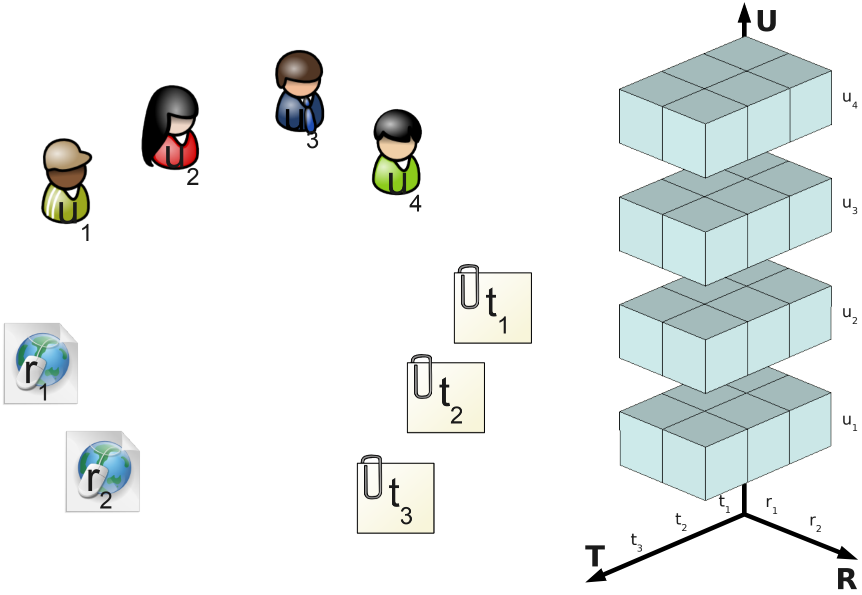
7. Triadic Formal Concept Analysis
   - Motivation
   - Folksonomies
   - Motivation
   - Triadic Formal Concept
   - Concept-Tri-Lattice
   - Visualization of Tri-Lattices
   - Iceberg Tri-Lattices
   - Computing Tri-Concepts
   - Qualitative Evaluation
   - Neighborhoods

Folksonomies

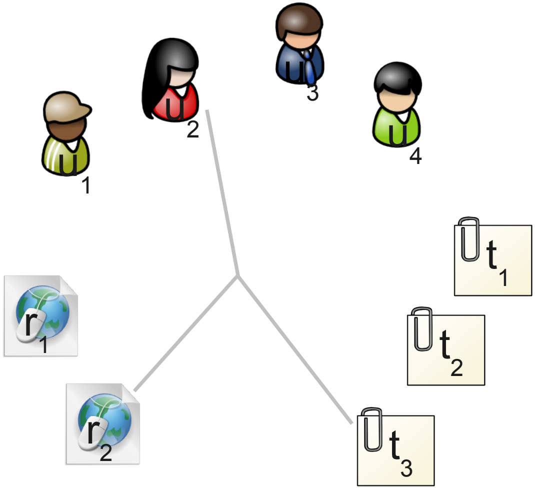
- data structure of collaborative tagging systems
- connects users, tags, and resources
- conceptual structure created by the people

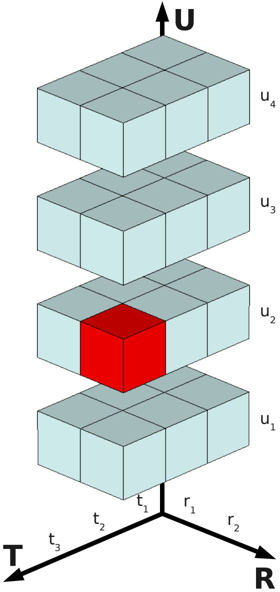


Folksonomies: Hypergraph, Tensor

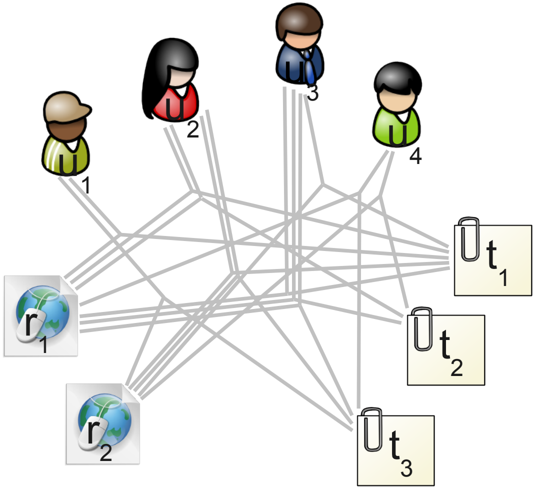


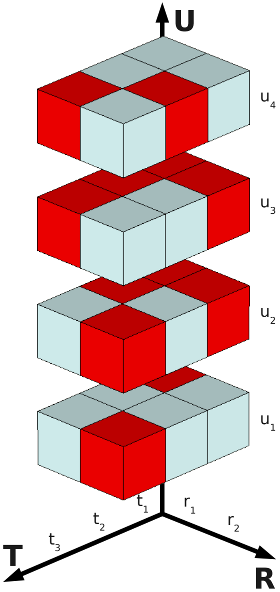
Folksonomies: Hypergraph, Tensor





Folksonomies: Hypergraph, Tensor





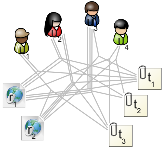
Folksonomies

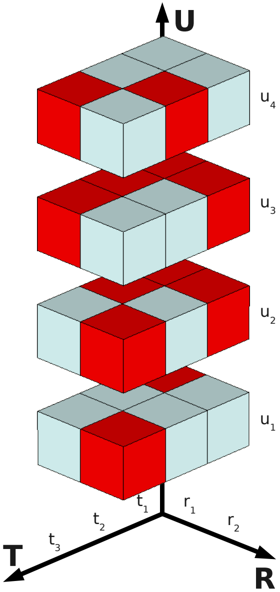
Definition (Folksonomy)

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

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

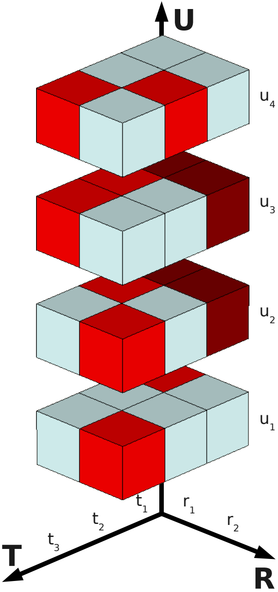
- tripartite hypergraph
- boolean 3-dimensional tensor
- triadic formal context





Motivation

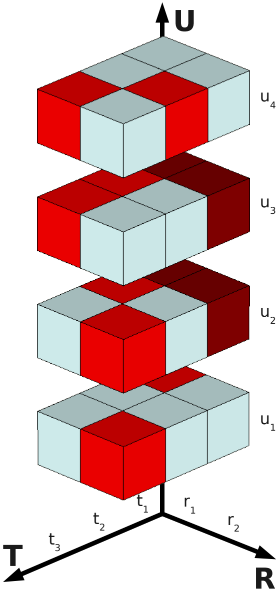
- conceptual clustering of folksonomies
  - find interessting concepts/clusters
  - support browsing, community detection, recommendations
  - get an overview into the structure of a folksonomy



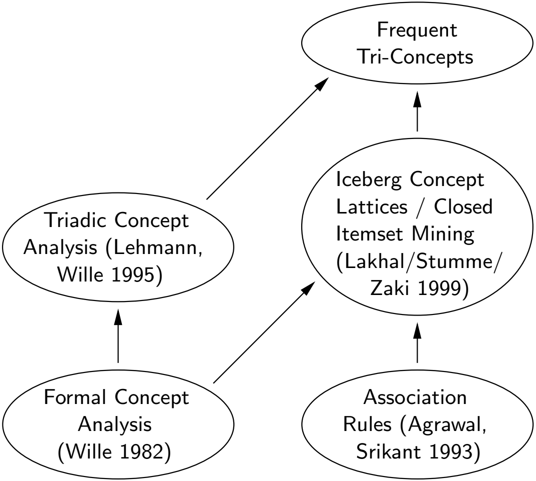
Motivation

- 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

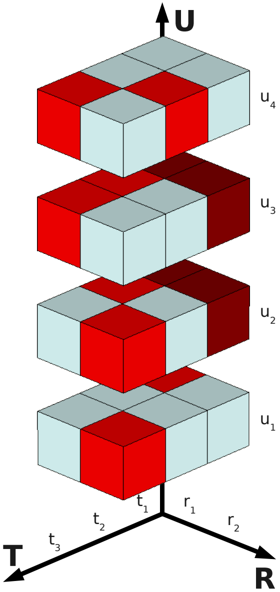


Motivation

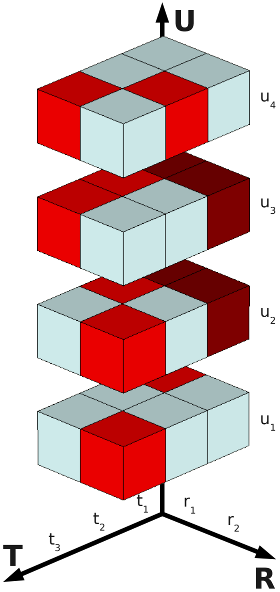

Frequent Tri-Concepts
Triadic Concept Analysis (Lehmann, Wille 1995)
Iceberg Concept Lattices / Closed Itemset Mining (Lakhal/Stumme/ Zaki 1999)
Formal Concept Analysis (Wille 1982)
Association Rules (Agrawal, Srikant 1993)

Motivation

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 Concept



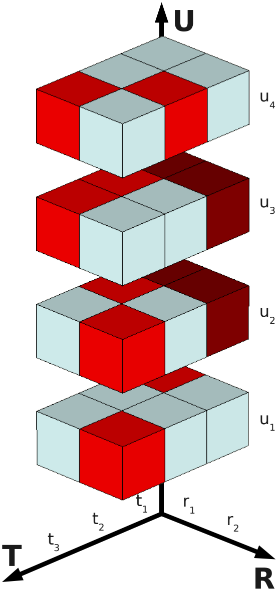
Definition (tri-concept)

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

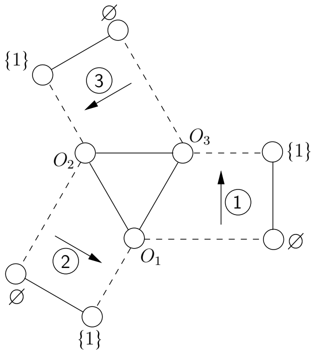
Concept-Tri-Lattice

- 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!

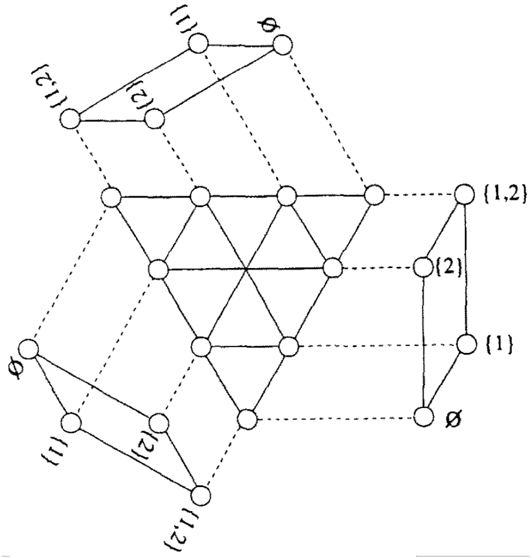


Visualization of Tri-Lattices

- 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 $\mathbb{B}_3 = \underline{\mathfrak{B}}(\{1\}, \{1\}, \{1\}, \emptyset)$
- visualization not always possible
  - satisfied tetrahedron condition
  - violated Thomson condition



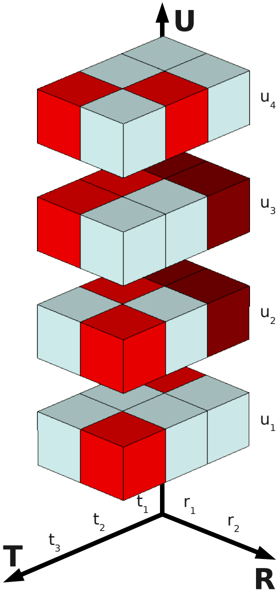
Visualization of Tri-Lattices



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 $|C| \geqslant \tau_r$

  → *iceberg tri-lattice*

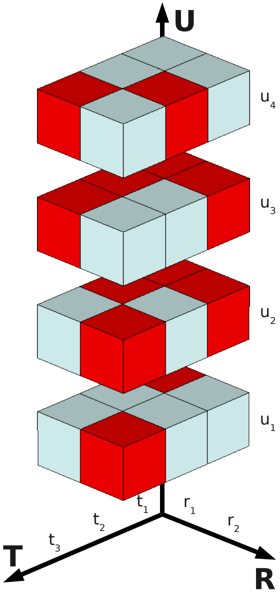


Computing Tri-Concepts

- 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

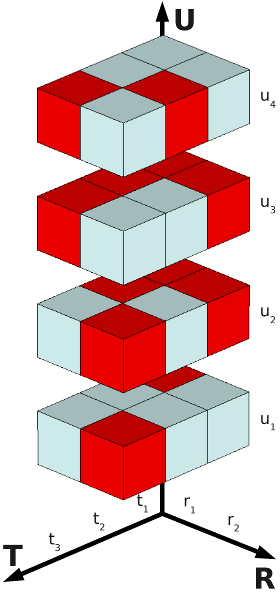
computes the iceberg tri-lattice of a triadic formal context



Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

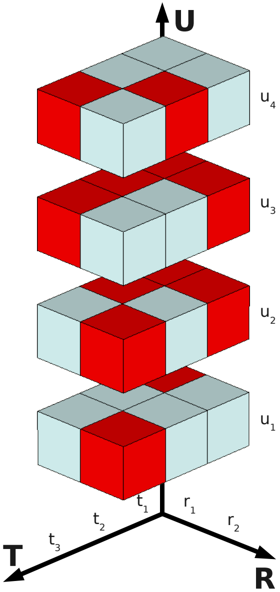


Computing Tri-Concepts

computes the iceberg tri-lattice of a triadic formal context

Algorithm

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

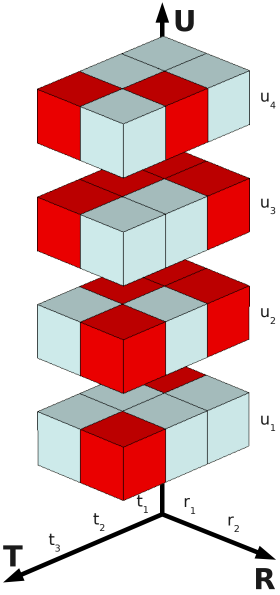


Computing Tri-Concepts

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

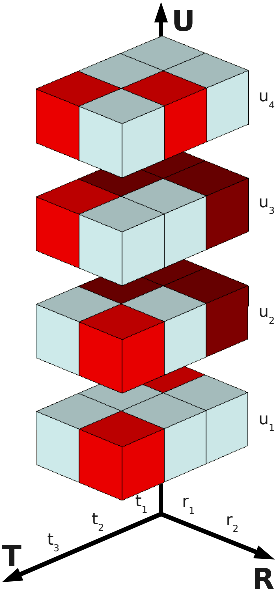
computes the iceberg tri-lattice of a triadic formal context

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:

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



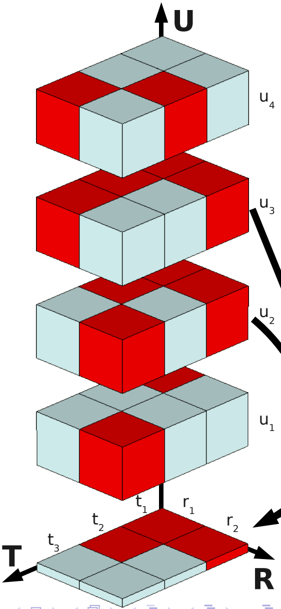
Computing Tri-Concepts

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



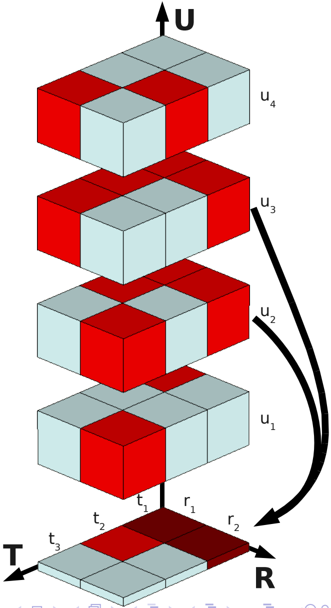
Computing Tri-Concepts

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

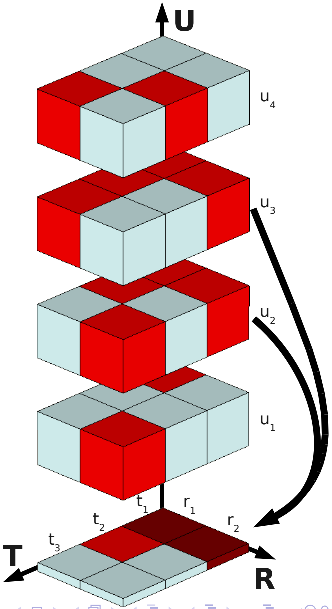
computes the iceberg tri-lattice of a triadic formal context

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

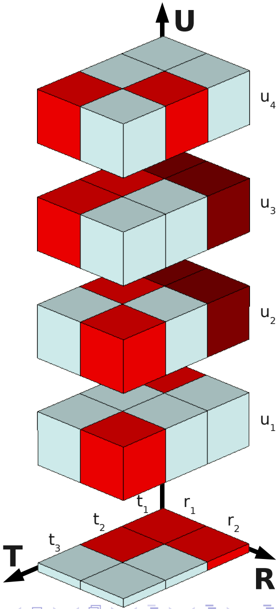
computes the iceberg tri-lattice of a triadic formal context

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:

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



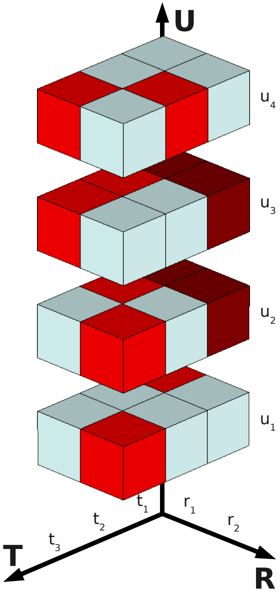
Computing Tri-Concepts

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

$A := \varnothing^I$

$B := A^I$

if $|A| < \tau$ **then**

  NextFrequentConcept$((A, B), (G, M, I), \tau)$

end if

return (A, B)

Computing Tri-Concepts

the *NextFrequentConcept* method:

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

i := *max*(M)
while *defined*(i) **do**
    $A := (B \bullet i)^I$
    if $|A| \geqslant \tau$ **then**
        $D := A^I$
        if $B <_i D$ **then**
            $B := D$
            return true
        end if
    end if
    $i :=$ *max*$(M \backslash B \cap \{1, \ldots, i-1\})$
end while
return false

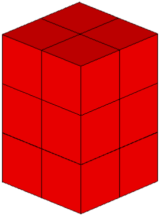
Qualitative Evaluation

BibSonomy Dataset:

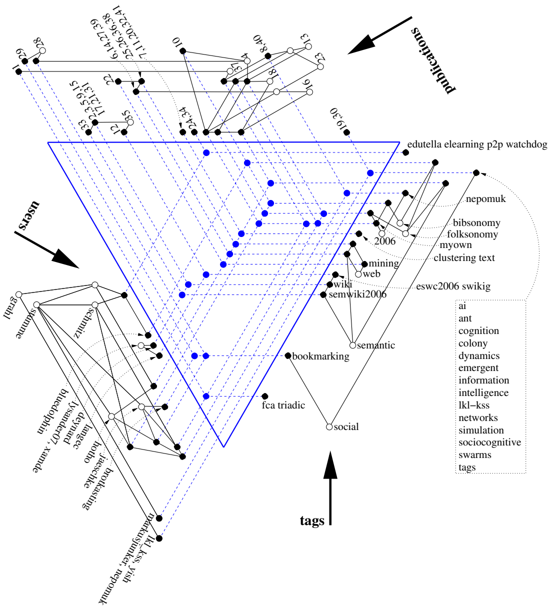
- all publication records until November 23rd, 2006
- removed: DBLP, posts with the tag “imported”
- $|U| = 262$, $|T| = 5\,954$, $|R| = 11\,101$, $|Y| = 44\,944$

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

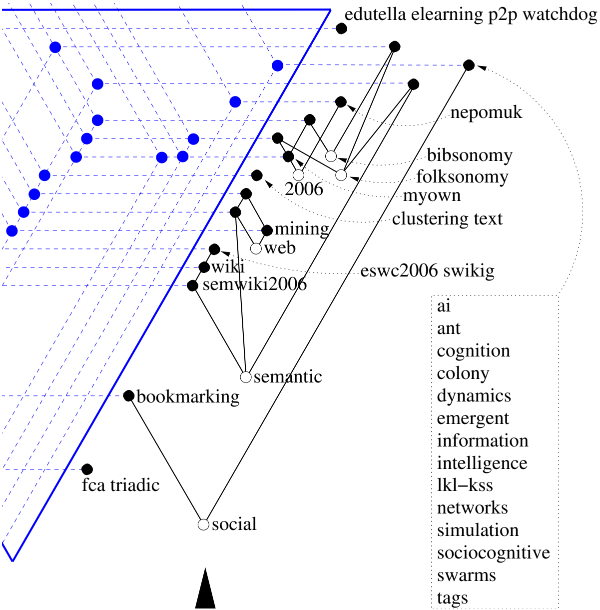


Qualitative Evaluation



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

Qualitative Evaluation



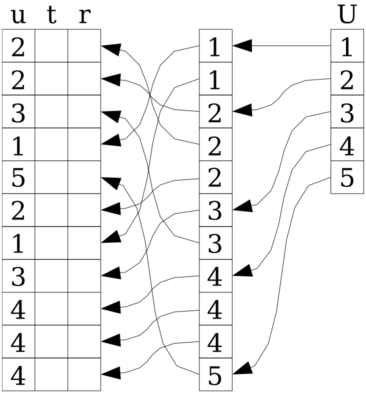
two topical groups:

- semantic
- social

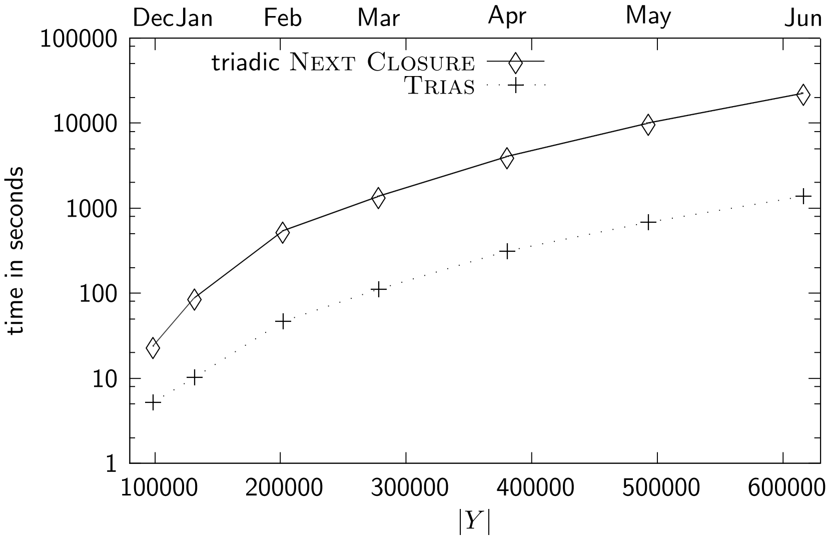
semantic further divided:

- wiki
- web
- folksonomy

Qualitative Evaluation



Qualitative Evaluation

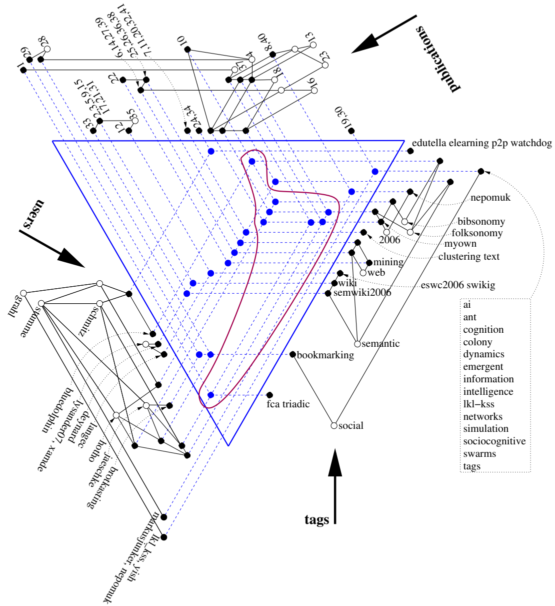


1	A Finite-State Model for On-Line Analytical Processing in Triadic Contexts
2	Annotation and Navigation in Semantic Wikis
3	A Semantic Wiki for Mathematical Knowledge Management
4	BibSonomy: A Social Bookmark and Publication Sharing System
5	Bringing the "Wiki-Way" to the Semantic Web with Rhizome
6	Building and Using the Semantic Web
7	Conceptual Clustering of Text Clusters
8	Content Aggregation on Knowledge Bases using Graph Clustering
9	Creating and using Semantic Web information with Makna
10	Emergent Semantics in BibSonomy
11	Explaining Text Clustering Results using Semantic Structures
12	Harvesting Wiki Consensus - Using Wikipedia Entries as Ontology Elements
13	Information Retrieval in Folksonomies: Search and Ranking
14	KAON – Towards a Large Scale Semantic Web
15	Kaukolu: Hub of the Semantic Corporate Intranet
16	Kollaboratives Wissensmanagement
17	Learning with Semantic Wikis
18	Mining Association Rules in Folksonomies
19	On Self-Regulated Swarms, Societal Memory, Speed and Dynamics
20	Ontologies improve text document clustering
21	Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics

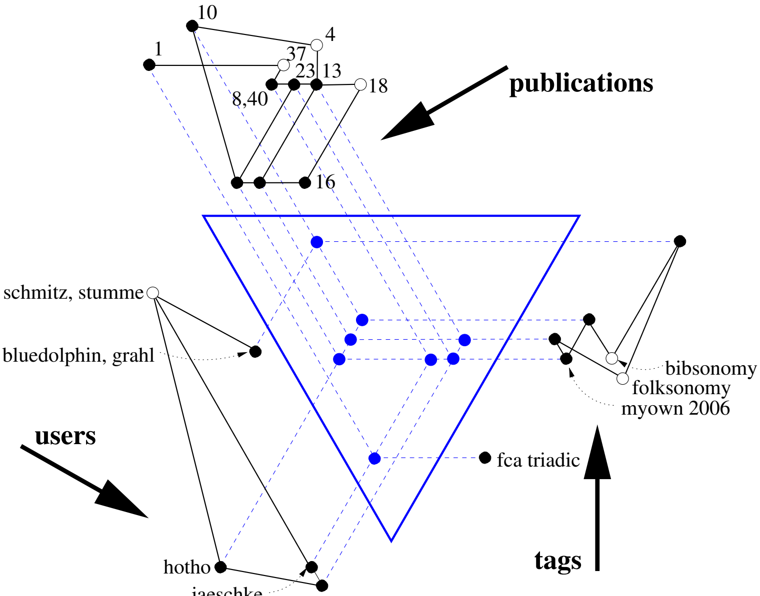
22	Proc. of the European Web Mining Forum 2005
23	Semantic Network Analysis of Ontologies
24	Semantic Resource Management for the Web: An ELearning Application.
25	Semantic Web Mining
26	Semantic Web Mining and the Representation, Analysis, and Evolution of Web Space
27	Semantic Web Mining for Building Information Portals (Position Paper)
28	Social Bookmarking Tools (I): A General Review
29	Social Bookmarking Tools (II). A Case Study – Connotea
30	Social Cognitive Maps, Swarm Collective Perception and Distributed Search on Dynamic Landscapes
31	SweetWiki : Semantic Web Enabled Technologies in Wiki
32	Text Clustering Based on Background Knowledge
33	The ABCDE Format Enabling Semantic Conference Proceedings
34	The Courseware Watchdog: an Ontology-based tool for Finding and Organizing Learning Material
35	Towards a Wiki Interchange Format (WIF) – Opening Semantic Wiki Content and Metadata
36	Towards Semantic Web Mining
37	TRIAS - An Algorithm for Mining Iceberg Tri-Lattices
38	Usage Mining for and on the Semantic Web (Book)
39	Usage Mining for and on the Semantic Web (Workshop)
40	Wege zur Entdeckung von Communities in Folksonomies
41	WordNet improves text document clustering

back

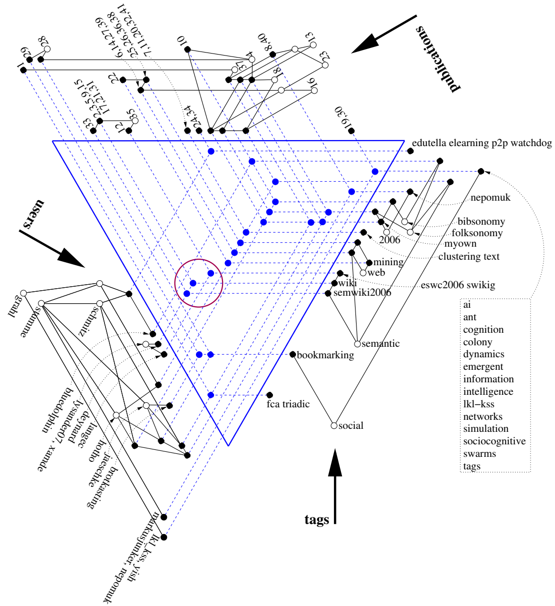
Qualitative Evaluation



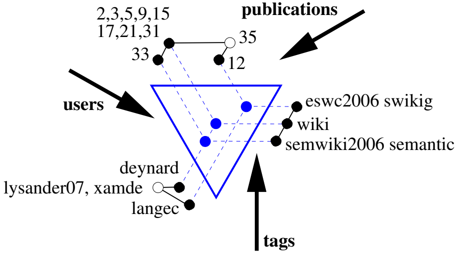
Qualitative Evaluation



Qualitative Evaluation



Qualitative Evaluation



publication 33: "'The ABCDE Format Enabling Semantic Conference Proceedings"'

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

$(\{jaeschke, schmitz, stumme\}, \{fca, triadic\}, \{1, 37\})$