1st Learning Analytics Summercamp

31 July 2015, Prague

Formal Concept Analysis and (Competence-based) Knowledge Space Theory

Michael Bedek
Knowledge Technologies Institute,
Graz University of Technology

LEA'S BOX
Learning Analytics Toolbox

TU Graz

Formal theories and methods

based on…

- Discrete mathematics
- Set- and Order Theory

for…

- Structuring and assigning *problems* / *test items*, *skills* / *competences* as well as learning objects / ressources

to enable …

- Adaptive knowledge and competence assessment
- Personalized competence development

Background and Motivation

Top-down, theory-driven, deductive

(Competence-based) Knowledge Space Theory

Knowledge Space Theory

Inductive Item Tree Analysis

Formal Concept Analysis

Bottom-up, data-driven, inductive

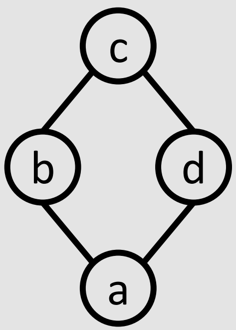
KST - Theoretical Foundations

Knowledge Space Theory (KST)

- Knowledge Domain Q consisting of a set of problems / items
- Prerequisite Relation $\preccurlyeq$ between items: $\preccurlyeq \subseteq Q \times Q$

$\preccurlyeq$	a	b	c	d
a	X	X	X	X
b		X	X	
c			X	
d			X	X

- reflexive
- antisymmetric
- transitive

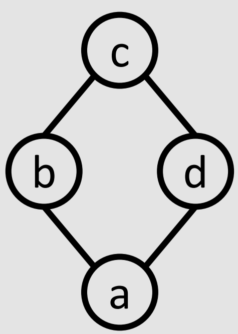


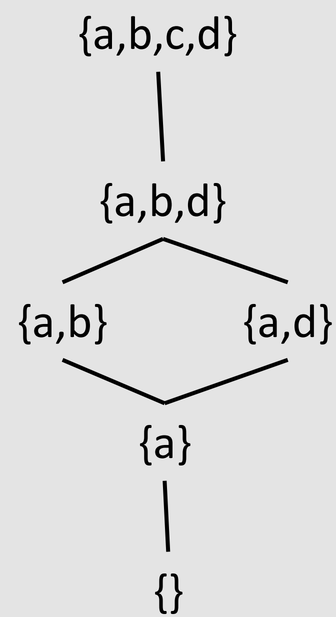
(Doignon & Falmagne, 1985, 1999)

KST - Theoretical Foundations

Knowledge Space Theory (KST)

- Knowledge Space is a pair $(Q,\mathcal{K})$
  - $\{\}, Q \in \mathcal{K}$
  - $\mathcal{K}$ is closed under union





(Cb)KST : Theoretical Foundations

Problem a : Compute the set of all divisors of 77.

D_P^- *Method 1:* (i) $77 = 1 \cdot 77 = 7 \cdot 11$
(ii) $D_{77} = \{1, 7, 11, 77\}$.

D_P^- *Method 2:* (i) $77 = 7 \cdot 11$
(ii) Divisors: $1, 7, 11, 77$, thus $D_{77} = \{1, 7, 11, 77\}$.

Problem b : Compute the set of all divisors of 230.

D_P *Method 1:* (i) $230 = 1 \cdot 230 = 2 \cdot 115 = 5 \cdot 46 = 10 \cdot 23$
(ii) $D_{230} = \{1, 2, 5, 10, 23, 46, 115, 230\}$.

D_P *Method 2:* (i) $230 = 2 \cdot 5 \cdot 23$
(ii) Divisors: $1, 2, 5, 23, 10, 46, 115, 230$;
thus $D_{230} = \{1, 2, 5, 10, 23, 46, 115, 230\}$.

Problem c : Compute the prime factorization of 273.

P *Method:* $273 = 3 \cdot 91 = 3 \cdot 7 \cdot 13$

Problem d : Compute the set of common divisors of 172 and 258.

C_D *Method 1:* (i) $D_{172} = \{1, 2, 4, 43, 86, 172\}$
$D_{258} = \{1, 2, 3, 6, 43, 86, 129, 258\}$
(ii) $D_{172} \cap D_{258} = \{1, 2, 43, 86\}$.

C_G *Method 2:* (i) $\mathrm{GCD}(172, 258) = 86$
(ii) $D_{172} \cap D_{258} = D_{86} = \{1, 2, 43, 86\}$.

Problem e : Compute the greatest common divisor of 275 and 385.

G_D *Method 1:* (i) $D_{275} \cap D_{385} = \{1, 5, 11, 55\}$
(ii) $\mathrm{GCD}(275, 385) = \max(D_{275} \cap D_{385}) = 55$.

G_P *Method 2:* (i) $275 = 5 \cdot 5 \cdot 11$; $385 = 5 \cdot 7 \cdot 11$
(ii) $\mathrm{GCD}(275, 385) = 5 \cdot 11 = 55$.

Problem f : Compute the least common multiple of 275 and 385.

L_P *Method 1:* (i) $275 = 5 \cdot 5 \cdot 11$; $385 = 5 \cdot 7 \cdot 11$
(ii) $\mathrm{LCM}(275, 385) = 5 \cdot 5 \cdot 7 \cdot 11 = 1925$.

L_G *Method 2:* (i) $\mathrm{GCD}(275, 385) = 55$
(ii) $\mathrm{LCM}(275, 385) = (275 \cdot 385)/55 = 1925$.

(Korossy, 1999)

(Cb)KST : Theoretical Foundations

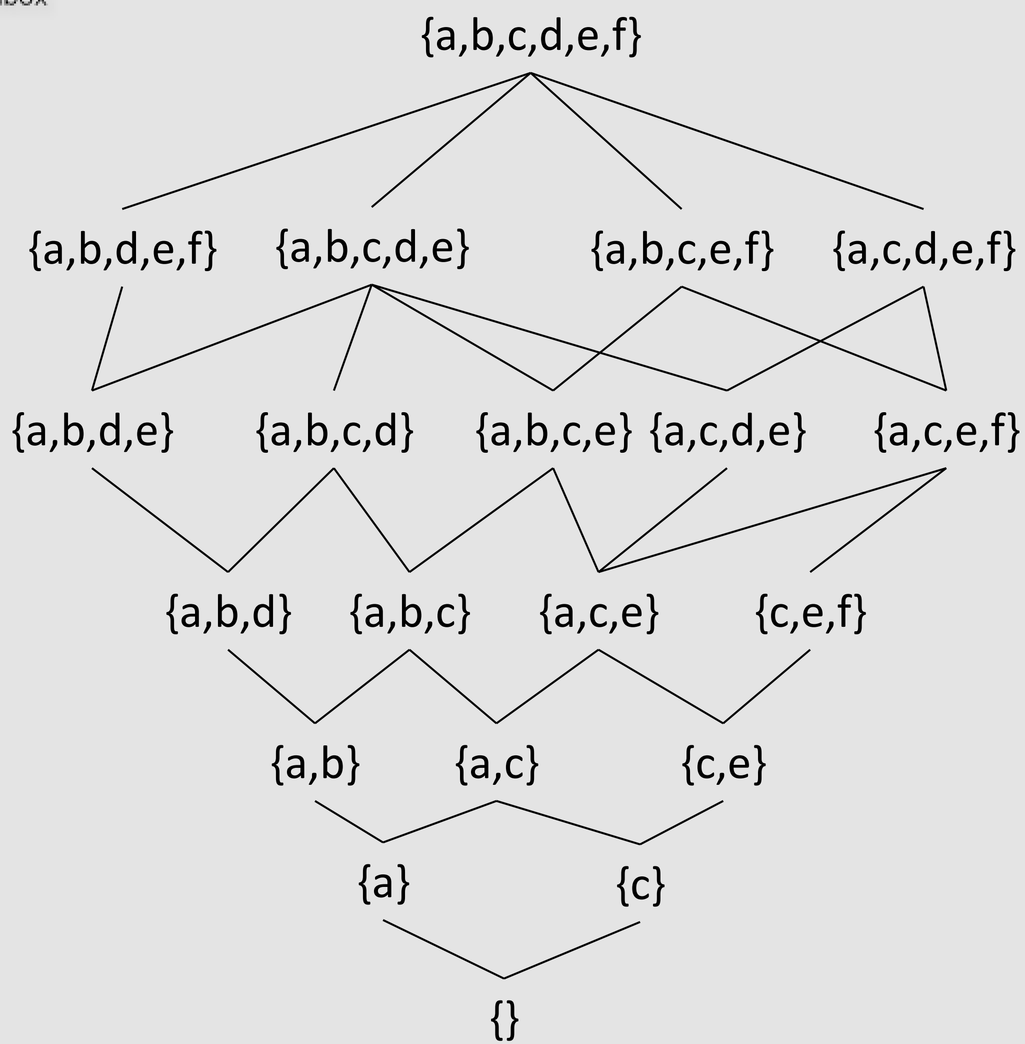
$\beta \in B(\mathcal{K})$	*Interpretation of the problems* a	b	c	d	e	f	$p(\beta)$
P			$\oplus$				c
D_p^-	$\oplus$						a
$D_{p'}^-$	$\oplus$						a
$PD_p^-D_p$	$+$	$\oplus$	$+$				abc
$D_{p'}^-D_{p'}$	$+$	$\oplus$					ab
$PD_p^-D_pC_D$	$+$	$+$	$+$	$\oplus$			$abcd$
$D_{p'}^-D_{p'}C_D$	$+$	$+$		$\oplus$			abd
$PD_p^-C_GG_p$	$+$		$+$	$\oplus$	$+$		$acde$
$PD_{p'}^-C_GG_p$	$+$		$+$	$\oplus$	$+$		$acde$
$PD_p^-D_pC_DG_D$	$+$	$+$	$+$	$+$	$\oplus$		$abcde$
$D_{p'}^-D_{p'}C_DG_D$	$+$	$+$		$+$	$\oplus$		$abde$
PG_p			$+$		$\oplus$		ce
PG_pL_p			$+$		$+$	$\oplus$	cef
PG_pL_G			$+$		$+$	$\oplus$	cef
$PD_p^-D_pC_DG_DL_G$	$+$	$+$	$+$	$+$	$+$	$\oplus$	$abcdef$
$D_{p'}^-D_{p'}C_DG_DL_G$	$+$	$+$		$+$	$+$	$\oplus$	$abdef$

(Cb)KST : Learner Modelling

$\beta \in B(\mathbb{K})$	*Interpretation of the problems* a	b	c	d	e	f	$p(\beta)$
P			$\circledast$				$\boxed{c}$
D_p^-	$\circledast$						$\boxed{a}$
$D_{\bar{p}}^-$	$\circledast$						a
$PD_p^-D_p$	$+$	$\circledast$	$+$				abc
$D_p^-D_p$	$+$	$\circledast$					$\boxed{ab}$
$PD_p^-D_pC_D$	$+$	$+$	$+$	$\circledast$			$abcd$
$D_p^-D_pC_D$	$+$	$+$		$\circledast$			$\boxed{abd}$
$PD_p^-C_GG_p$	$+$		$+$	$\circledast$	$+$		$\boxed{acde}$
$PD_{\bar{p}}^-C_GG_p$	$+$		$+$	$\circledast$	$+$		$acde$
$PD_p^-D_pC_DG_D$	$+$	$+$	$+$	$+$	$\circledast$		$abcde$
$D_p^-D_pC_DG_D$	$+$	$+$		$+$	$\circledast$		$\boxed{abde}$
PG_p			$+$		$\circledast$		$\boxed{ce}$
PG_pL_p			$+$		$+$	$\circledast$	$\boxed{cef}$
PG_pL_G			$+$		$+$	$\circledast$	cef
$PD_p^-D_pC_DG_DL_G$	$+$	$+$	$+$	$+$	$+$	$\circledast$	$abcdef$
$D_p^-D_pC_DG_DL_G$	$+$	$+$		$+$	$+$	$\circledast$	$\boxed{abdef}$

$\mathcal{B}(K)$ =
{a, c, ab, abd, acde,
abde, ce, cef, abdef}

(Cb)KST : Learner Modelling



LEA´S BOX
Learning Analytics Toolbox

IITA : Theoretical Foundations

	a	b	c	d	e	f
01	1	0	1	1	1	1
02	1	1	1	0	1	1
03	1	1	1	1	1	1
04	1	1	0	0	0	0
05	1	1	1	1	1	0
06	1	0	1	0	0	0
07	0	1	0	1	1	1
08	1	0	1	0	0	0
09	1	1	1	0	0	0
10	0	0	1	0	0	0
11	1	1	1	0	1	0
12	1	1	1	0	1	0

	a	b	c	d	e	f
13	1	0	1	0	1	1
14	1	1	1	0	0	0
15	1	1	1	0	0	1
16	1	0	1	0	0	0
17	1	1	1	1	1	1
18	1	0	1	0	0	0
19	1	0	1	0	0	0
20	1	1	1	0	0	0
21	1	1	1	0	0	0
22	1	0	1	0	0	0
23	1	0	1	0	1	1

(van Leeuwe, 1974; data from Korossy, 1999)

IITA: Theoretical Foundations

j

i

bij	a	b	c	d	e	f
a	0	1	1	1	1	1
b	9	0	10	1	3	3
c	1	2	0	1	1	1
d	17	9	17	0	5	4
e	12	8	10	0	0	1
f	14	8	14	1	3	0

1.) bij – values

bij : „Violations“ (Counterexamples) of $j \mapsto i$ (or $i \leq j$)

$$b_{ij} := | \{ d \in D \mid d(j) = 1 \wedge d(i) = 0 \} |$$

2.) Stepwise inclusion of item-pairs with low bij – values to ≤ - i.e. the Quasiorder (M, ≤)

3.) Calculating reproducibility coefficient at each step – i.e. for each Quaisorder $(M, \leq_L)$

(Schrepp, 2003, 2006)

IITA : Application

	j						
	bij	a	b	c	d	e	f
	a	0	1	1	1	1	1
	b	9	0	10	1	3	3
i	c	1	2	0	1	1	1
	d	17	9	17	0	5	4
	e	12	8	10	0	0	1
	f	14	8	14	1	3	0

$b \longrightarrow a$

$c \longrightarrow a$

$d \longrightarrow a$

$e \longrightarrow a$

$f \longrightarrow a$

$d \longrightarrow b$

$a \longrightarrow c$

$b \longrightarrow c$

$d \longrightarrow c$

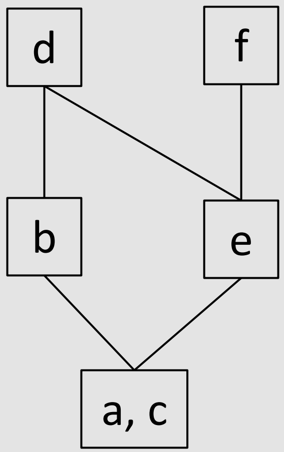
$e \longrightarrow c$

$f \longrightarrow c$

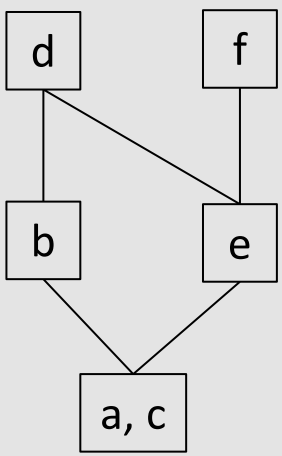
$d \longrightarrow e$

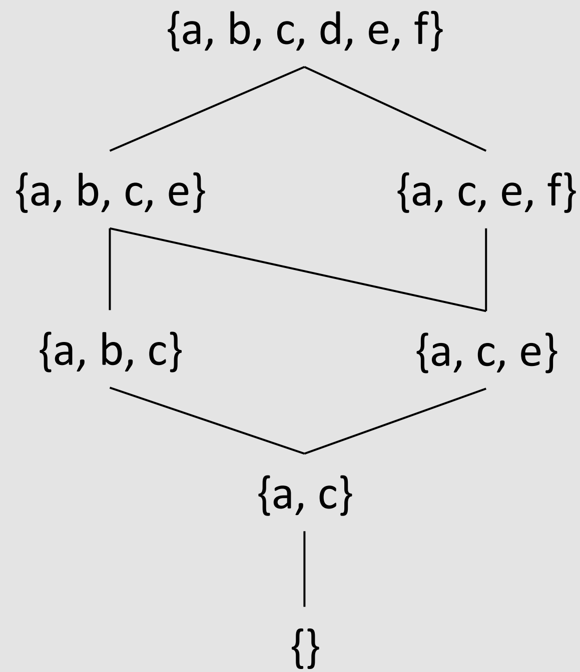
$f \longrightarrow e$

$d \longrightarrow f$



IITA : Theoretical Foundations





EA´S BOX Learning Analytics Toolbox

FCA: Theoretical Foundations

- Set of Objects G
- Set of Attributes M
- Binary Relation $I \subseteq G \times M$

$g \in G$ and $m \in M$

$g\,I\,m$:= object g has attribute m

- *Formal Context* K is a set structure $K := (G, M, I)$

(Wille 1982, 2005)

Definition of the formal context

	is toxic	hatched from egg	is able to fly	lives in/ on the water	is able to swim	live birth	performs photosynt...	bear fruits
Bumble-bee	✓	✓	✓					
Bee	✓	✓	✓					
Tree frog		✓		✓	✓			
Goldfish				✓	✓			
Root vole						✓		

FCA : Theoretical Foundations

$A \in G$ and $B \in M$, the following derivation operators need to be defined:

$A \mapsto A' := \{m \in M \mid g\, I\, m \text{ for all } g \in A\}$, which is the set of common attributes of the objects in A, and
$B \mapsto B' := \{g \in G \mid g\, I\, m \text{ for all } m \in B\}$, which is the set of objects which have all attributes of B in common.

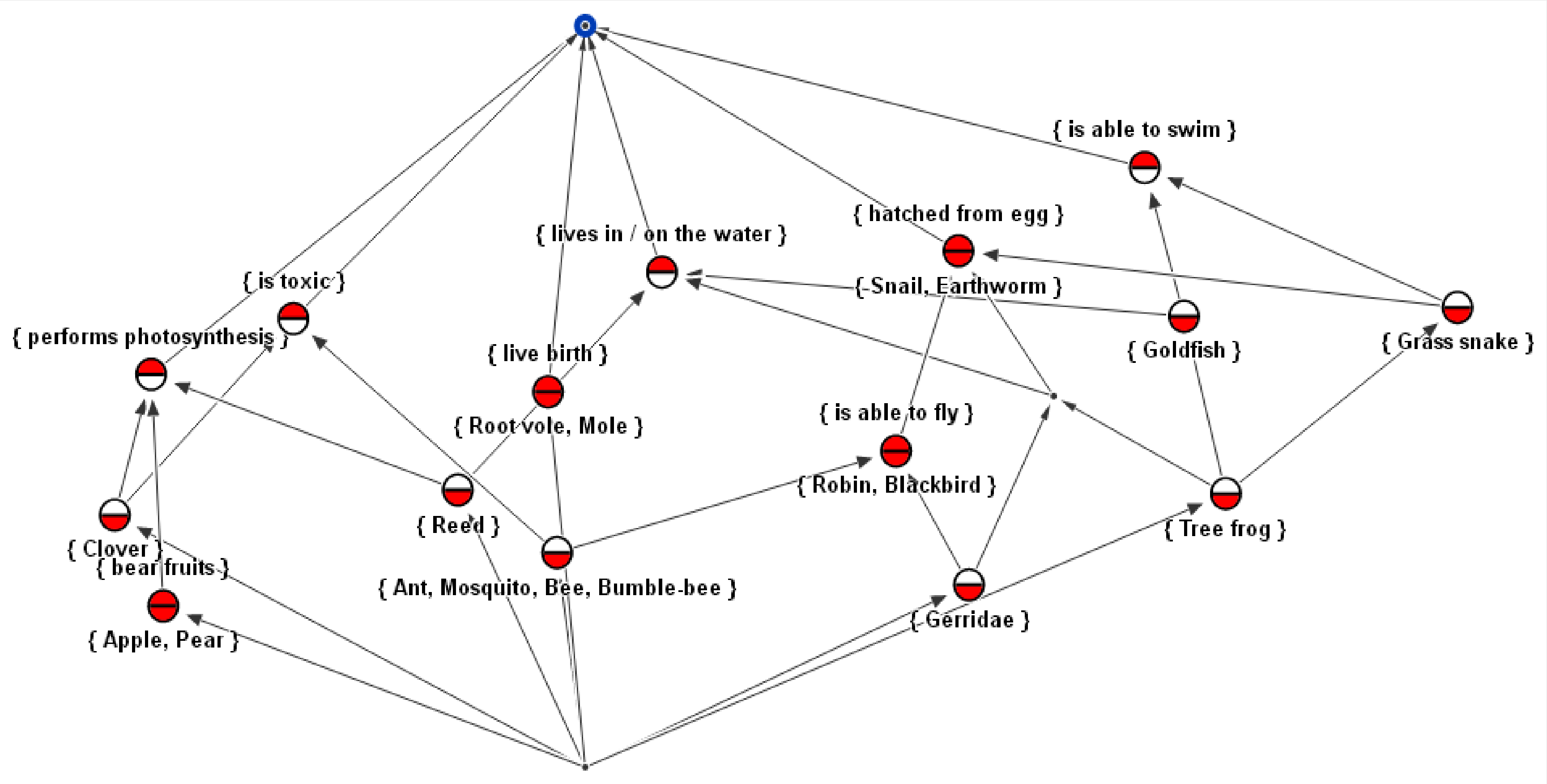
A formal concept is a pair (A, B) with the subsets $A \subseteq G$ and $B \subseteq M$ which fulfill

$A' = B$ and $B' = A$.

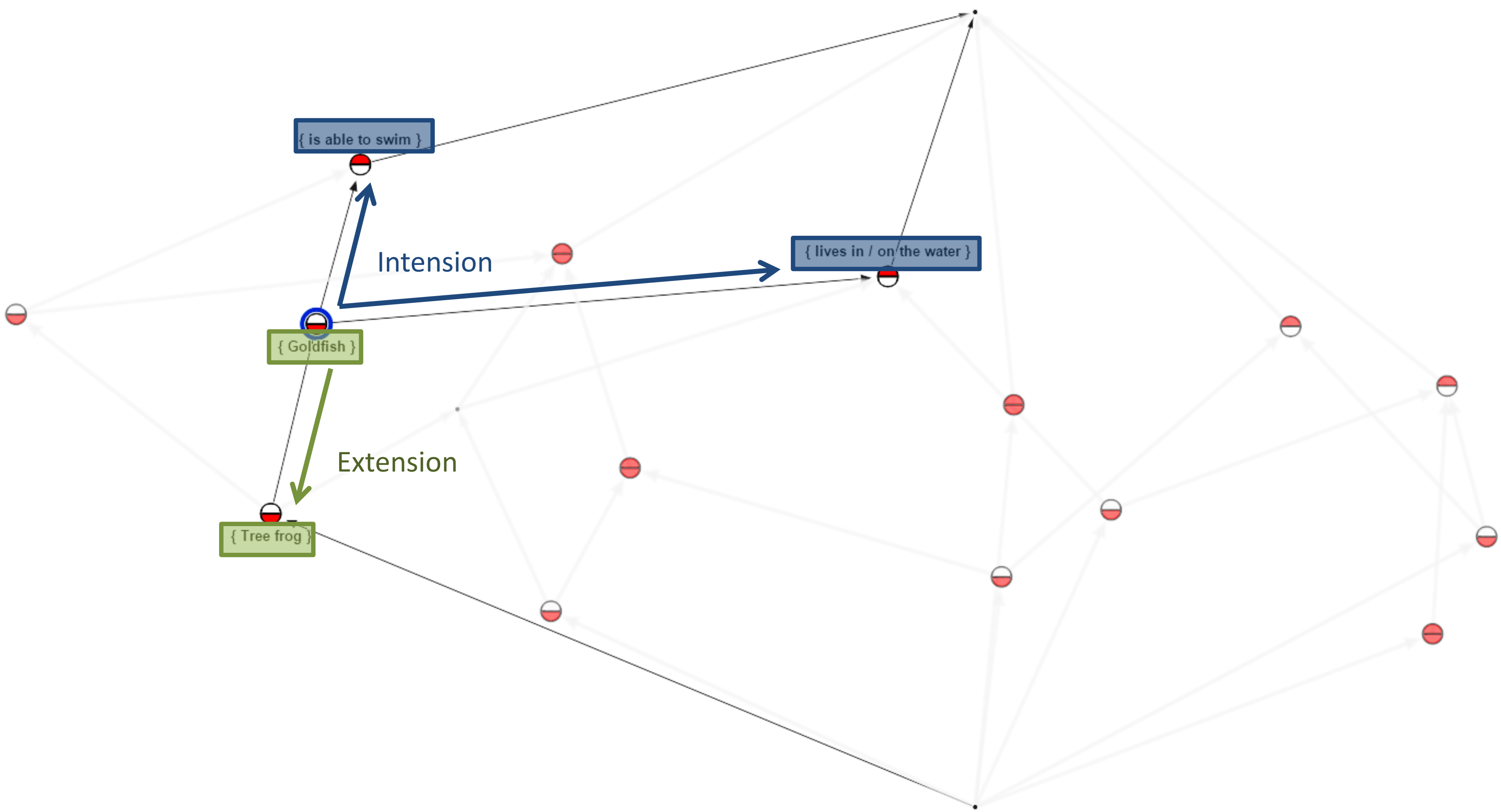
A = "*extension*", B = "*intension*"

$\mathcal{B}$(K) = Concept Lattice

Concept Lattice



“Reading” the Concept Lattice



Applying the FCA: Learner Modelling

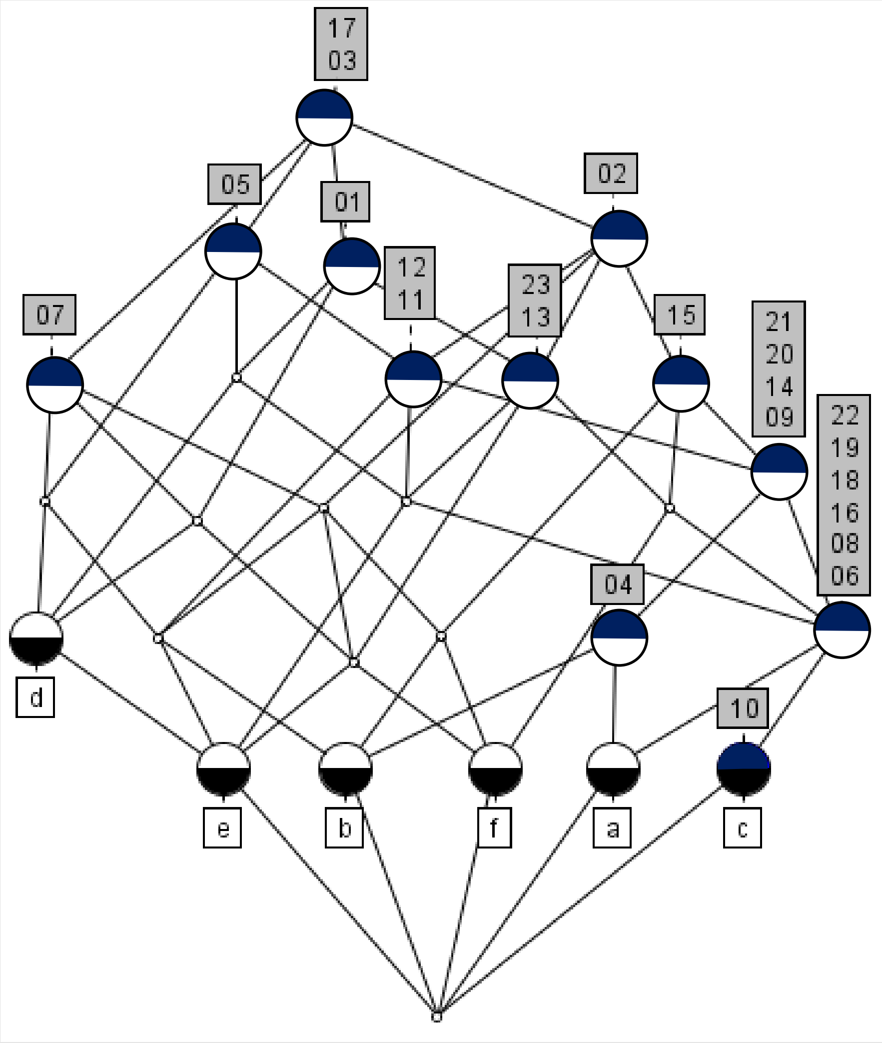
Definition of the "knowledge context"

(Rusch & Wille, 1996)

Problem \ Students	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
a	X	X	X	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
b		X	X	X	X		X		X		X	X		X	X		X			X	X		
c	X	X	X		X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
d	X		X		X		X										X						
e	X	X	X		X		X				X	X	X				X						X
f	X	X	X				X						X		X		X						X

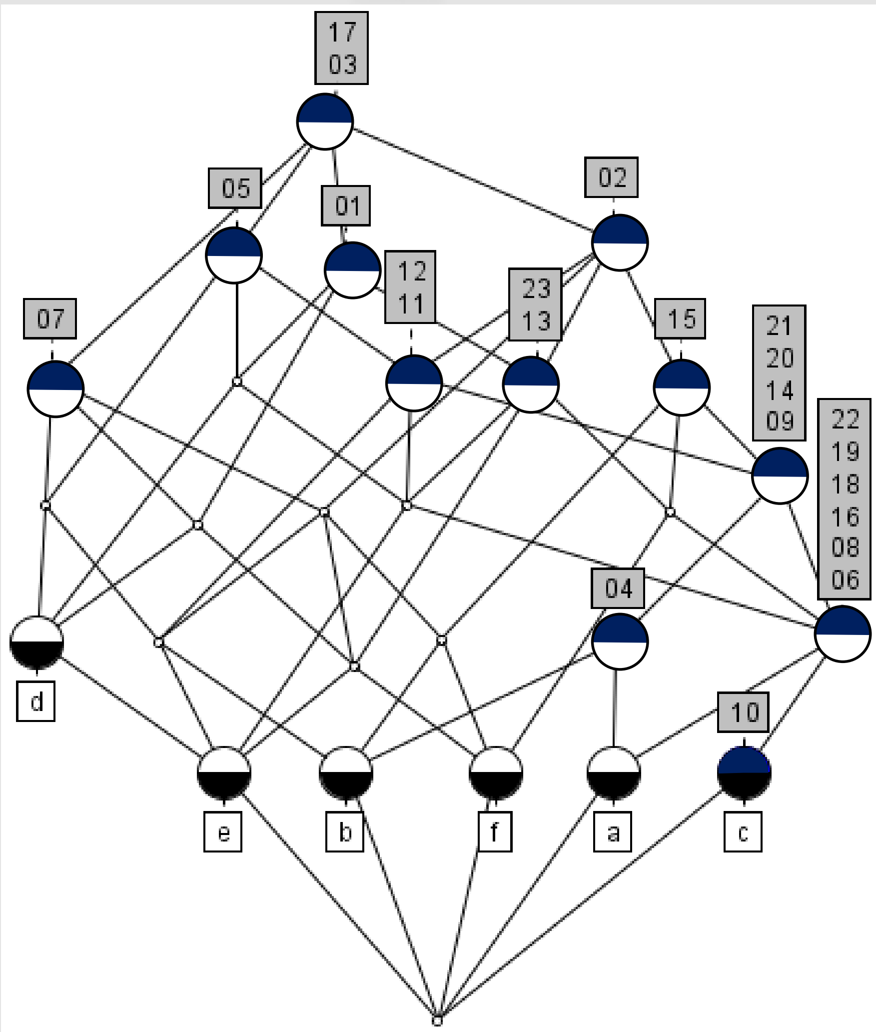
(data from Korossy, 1999)

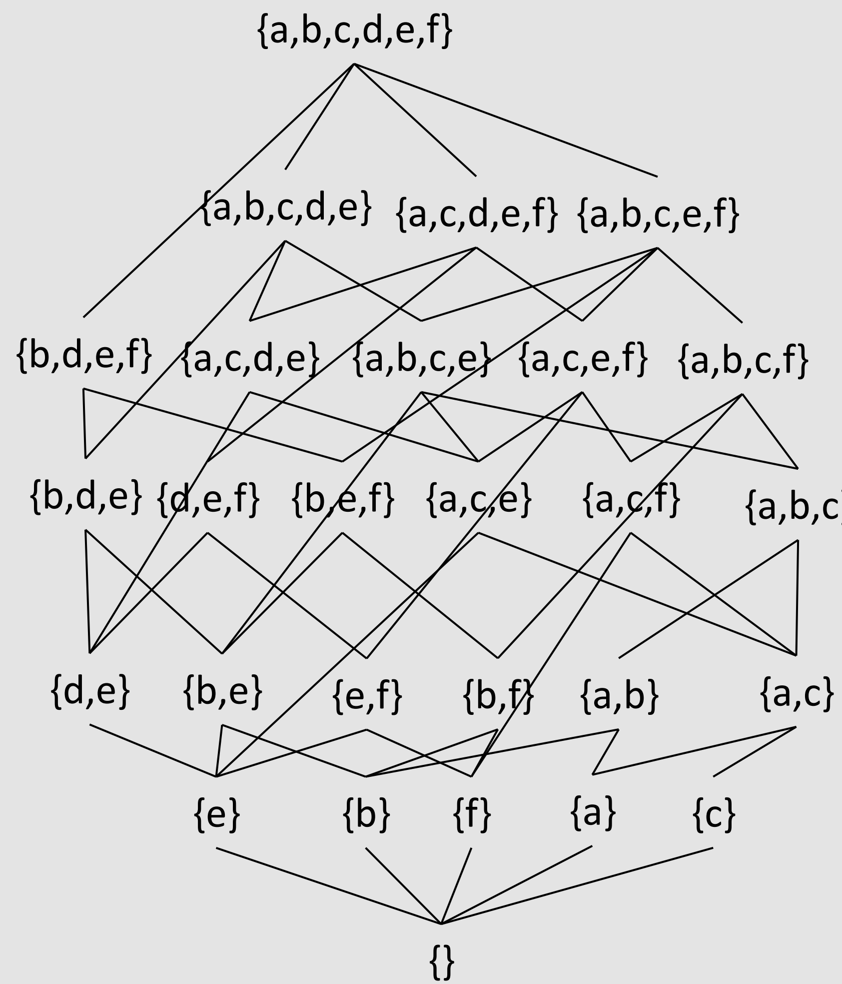
FCA : Learner Modelling



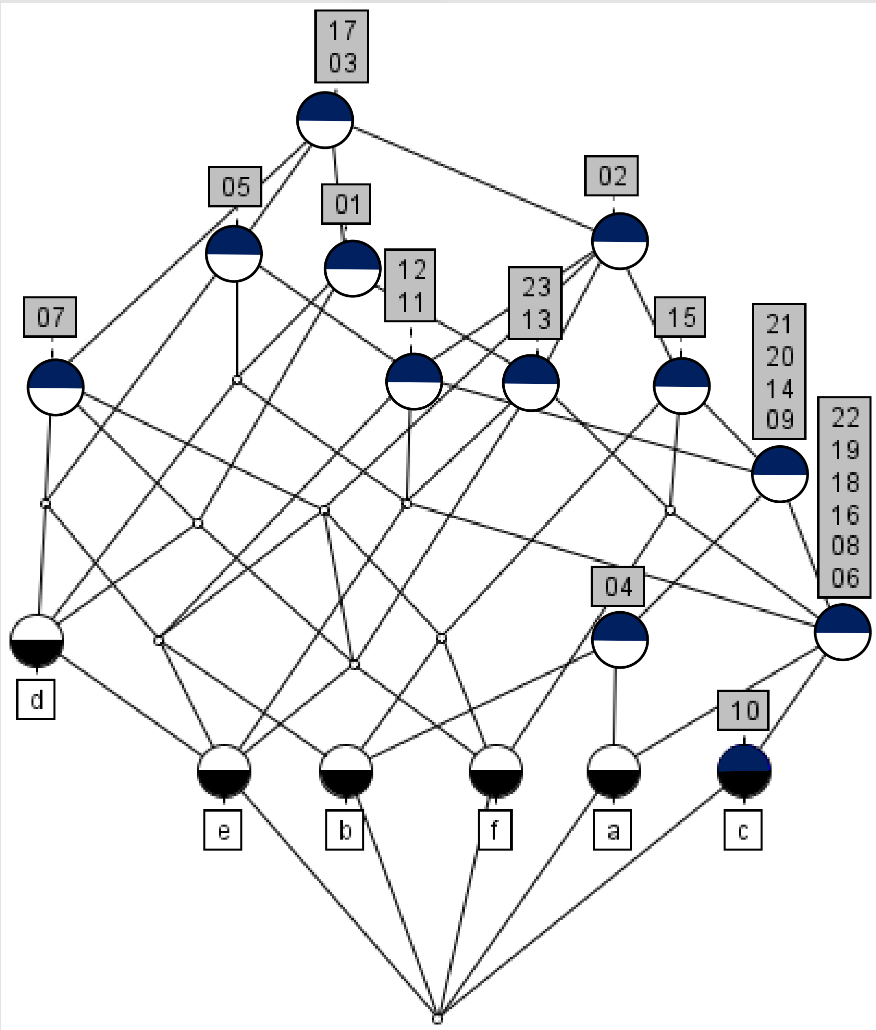
LEA'S BOX Learning Analytics Toolbox

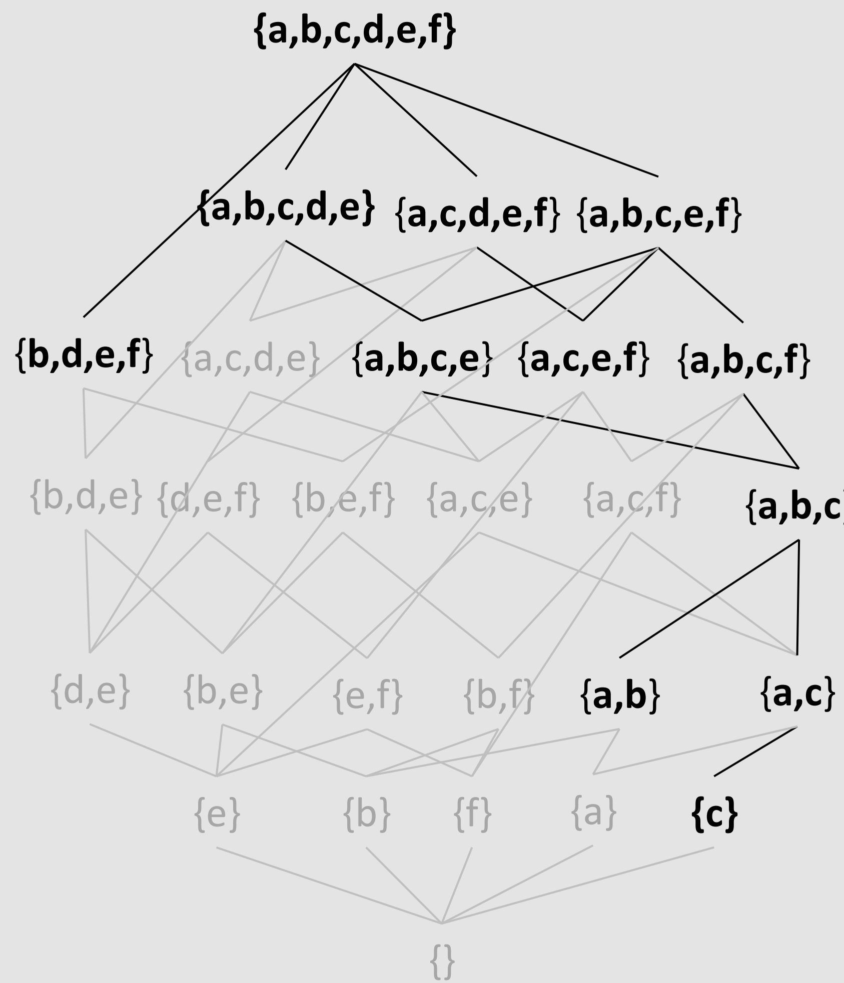
FCA : Learner Modelling





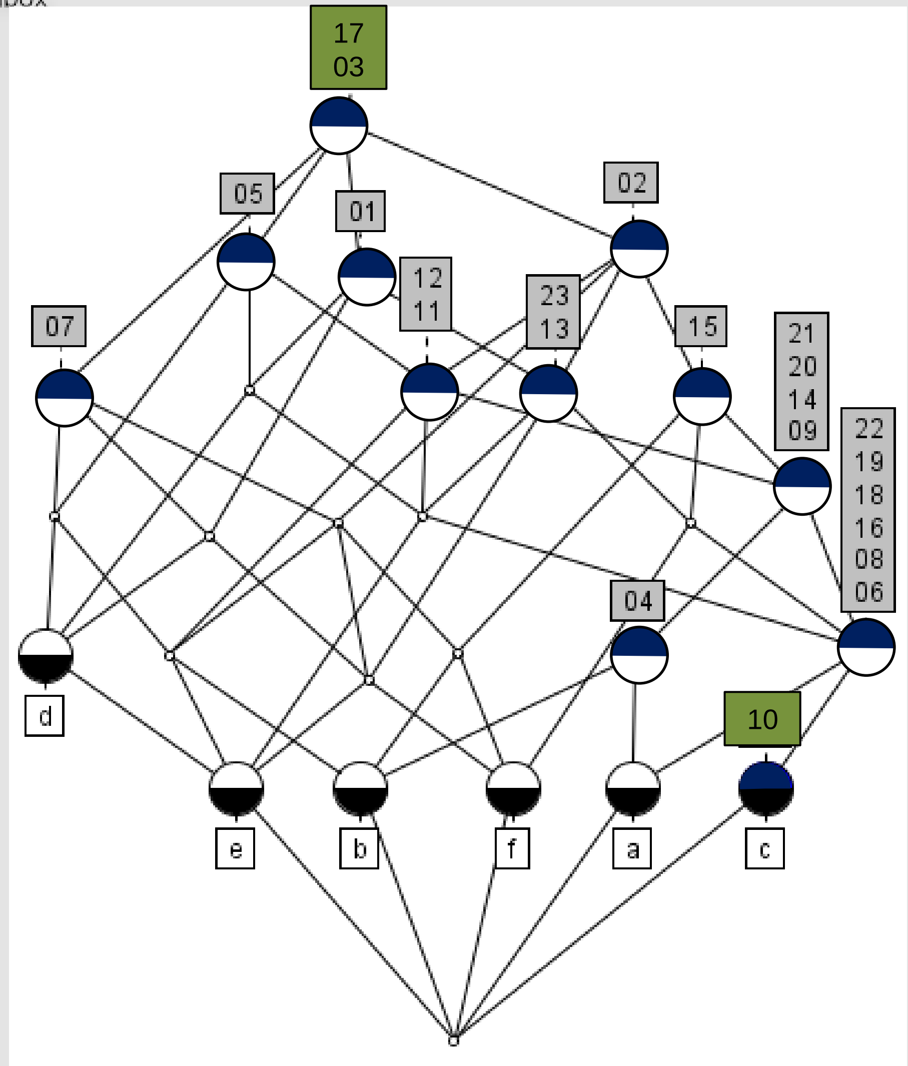
FCA : Learner Modelling





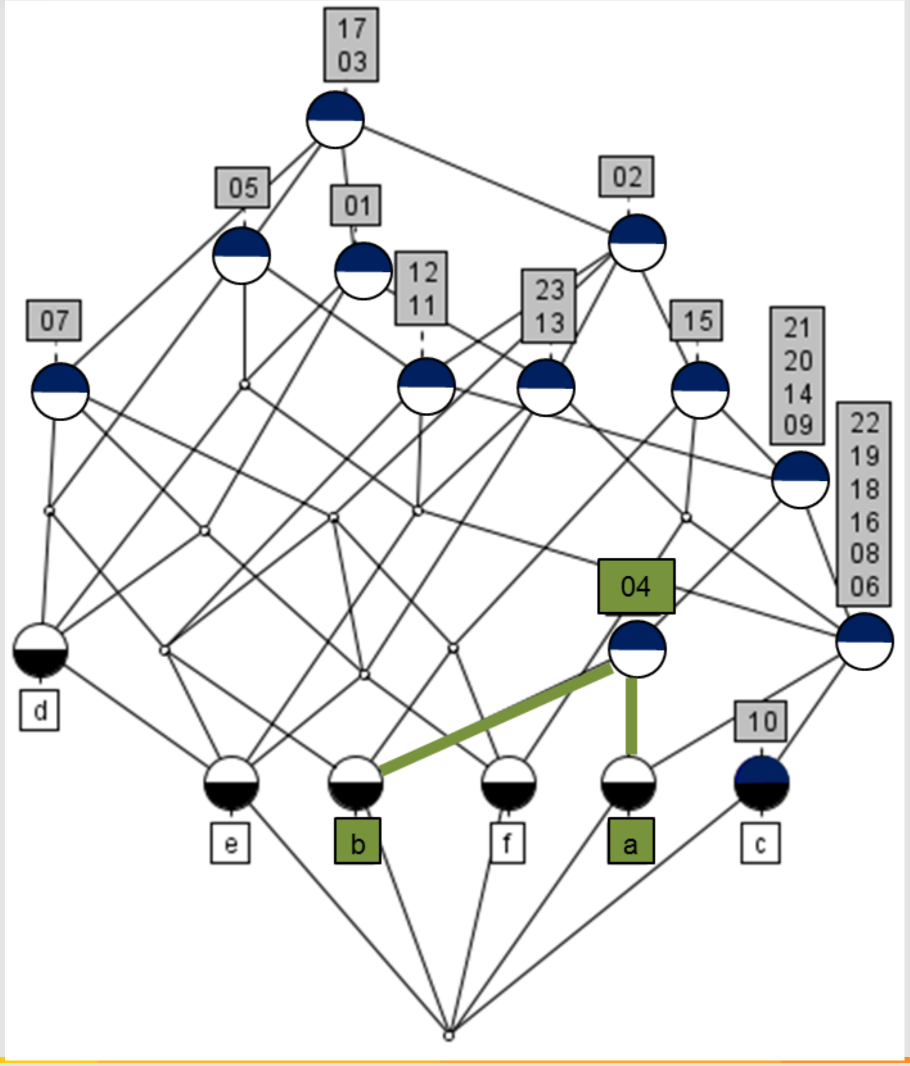
FCA : Learner Modelling

Who are the best / weakest students?



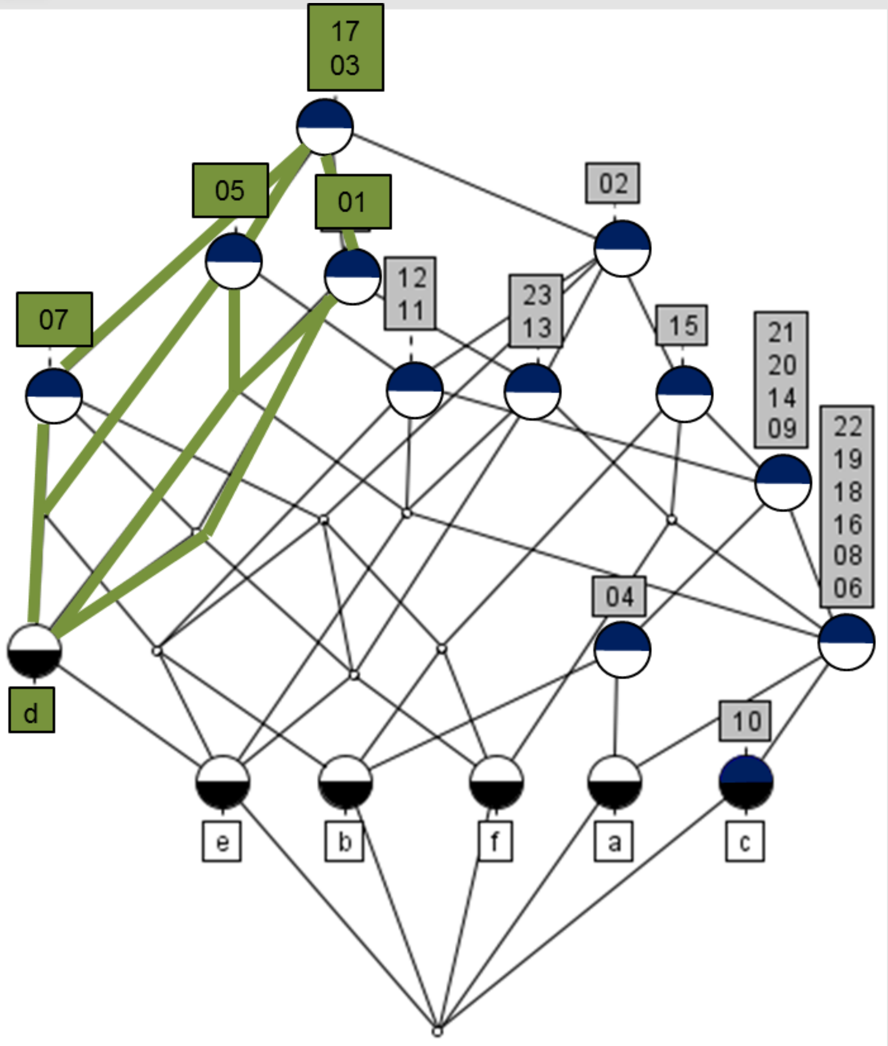
FCA : Learner Modelling

Which set of problems have been solved by a (set of) student(s)?



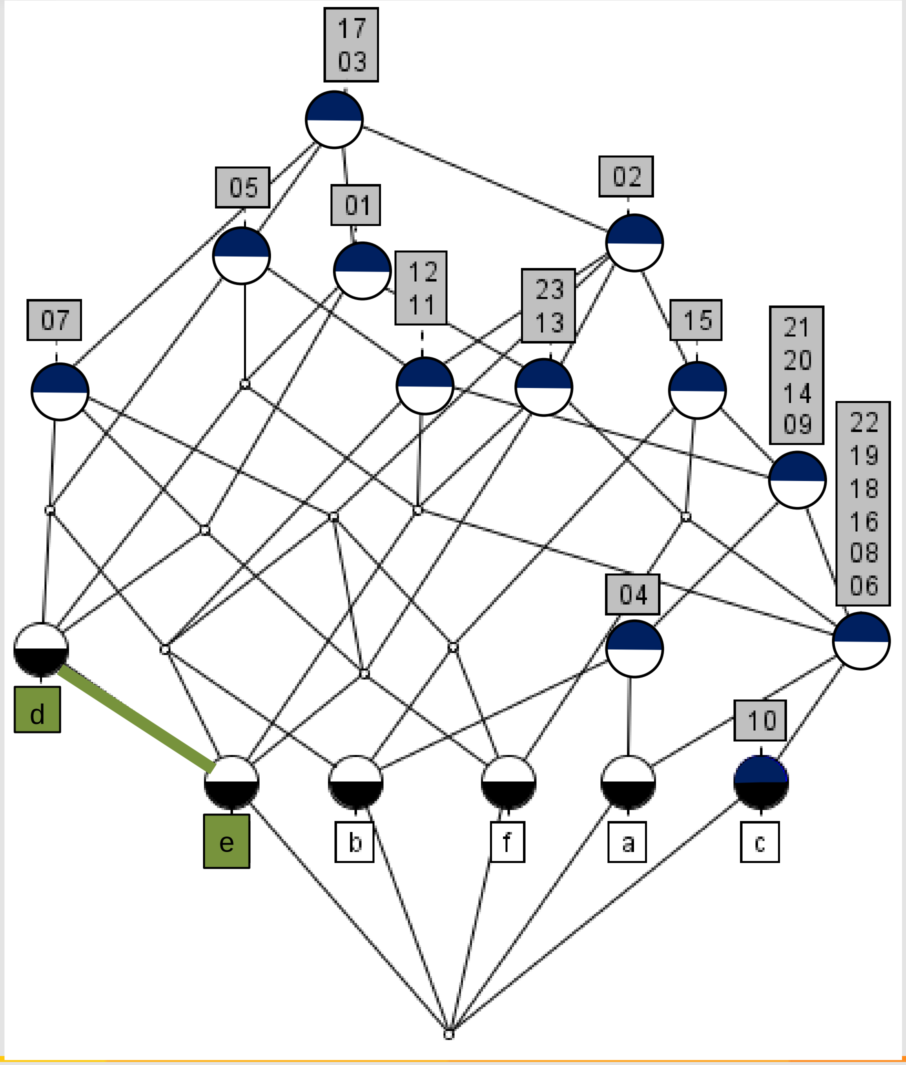
FCA : Learner Modelling

Whose students have solved a particular (set of) problem(s)?



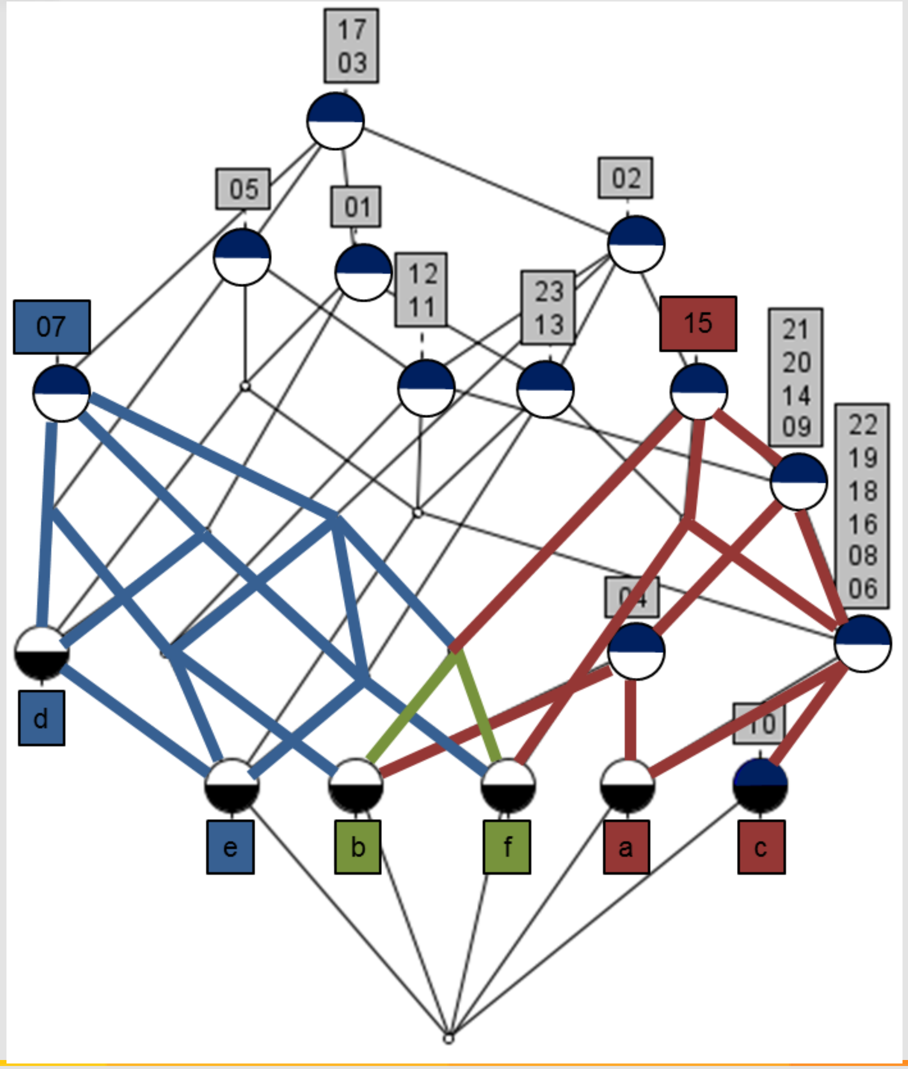
FCA : Learner Modelling

Are there surmise / prerequisite - relations between problems?

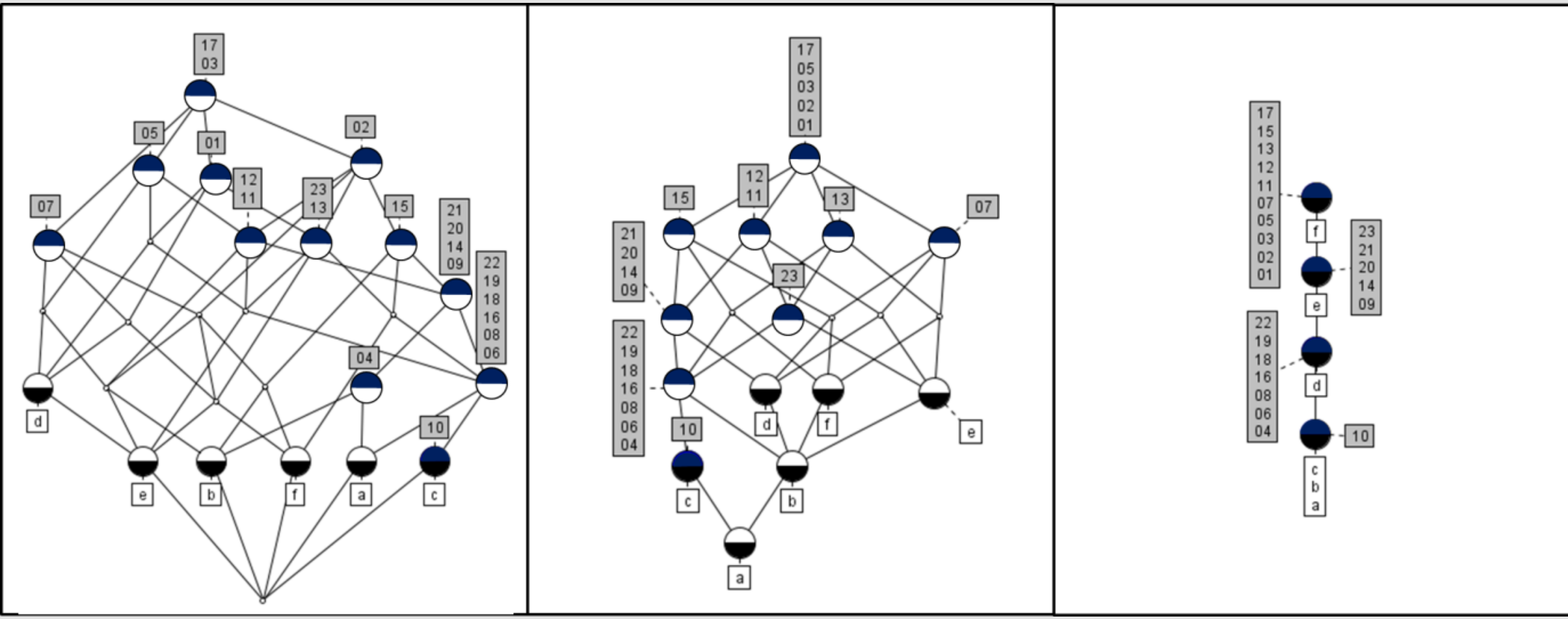


FCA : Learner Modelling

Comparing (sets) of students…

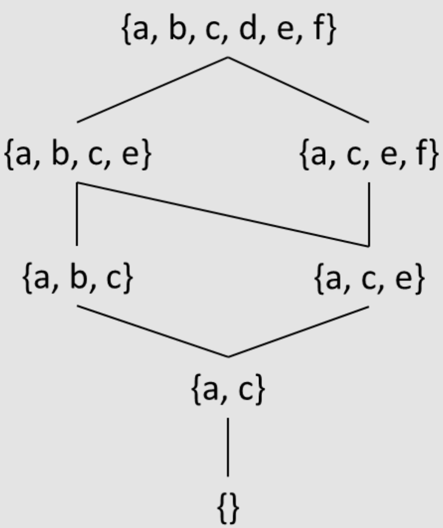


FCA : Learner Modelling

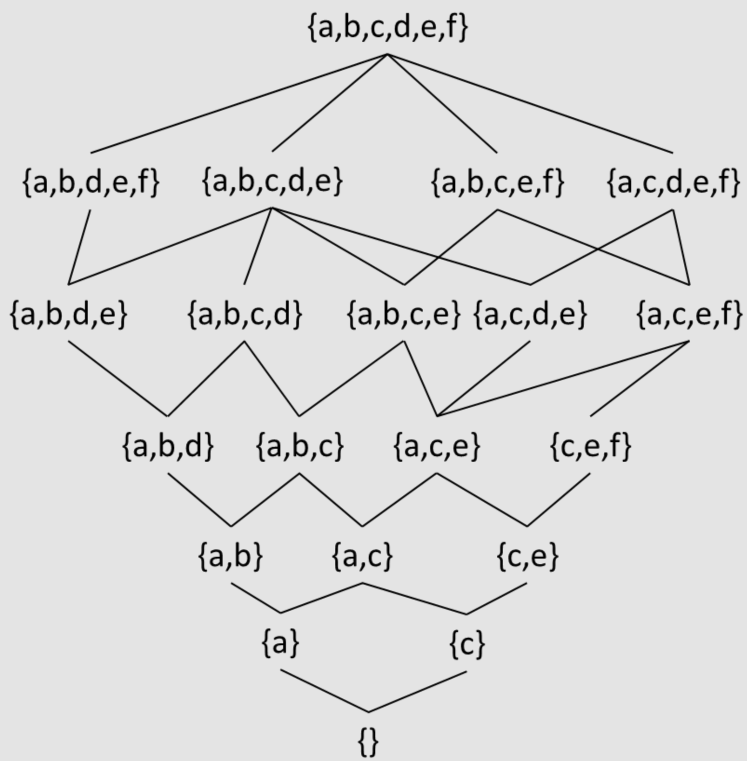


Visualizing learning progress on a classroom-level…

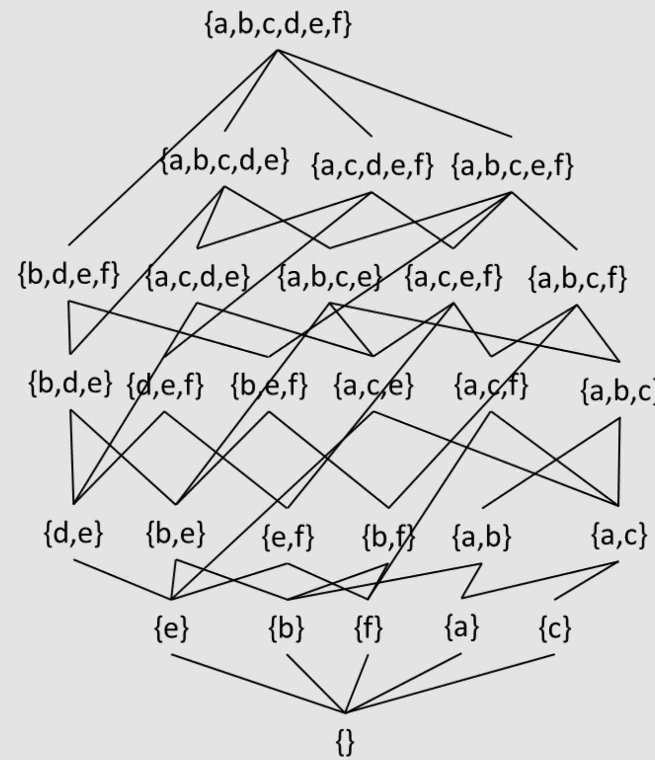
Comparisons



IITA



KST



FCA

LEA´S BOX
Learning Analytics Toolbox

Discussion & Current / Future Work

- FCA enables to answer a broad set of pedagogically relevant questions
- Vulnerability and lack of robustness due to careless errors and lucky quesses
- Exploiting and combining strengths of different methods:
  - IITA for establishing initial knowledge space,
  - KST for knowledge assessment,
  - FCA for visualizing students performances
- Simplified FCA visualizations without loss of information
- Extending with skills / competences and learning objects / resources

References

Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*, 175–196.

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge Spaces*. Berlin: Springer.

Korossy, K. (1999). Modeling Knowledge as Competence and Performance. In D. Albert & J. Lukas (Eds.), *Knowledge Spaces: Theories, Empirical Research, and Applications* (pp. 103-132). Mahwah, NJ: Springer.

Rusch, A., & Wille, R. (1996). Knowledge spaces and formal concept analysis. In H. H. Bock & W. Polasek (Eds.), *Data analysis and information systems: Statistical and conceptual approaches* (pp. 427-436). Berlin: Springer

Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire. *Methods of Psychological Research*, *19*, 43-79.

Schrepp, M. (2006). ITA 2.0: A Program for Classical and Inductive Item Tree Analysis. *Journal of Statistical Software*, *16*(10), 1-14.

Wille, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered sets* (pp. 445–470). Dordrecht: Reidel.

Wille, R. (2005). Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies, In B. Ganter , G. Stumme and R. Wille (Eds.), *Formal Concept Analysis* (pp. 1-34). Berlin: Springer.

Van Leeuwe, J.F.J. (1974). Item tree analysis. *Nederlands Tijdschrift voor de Psychologie*, *29*, 475-484