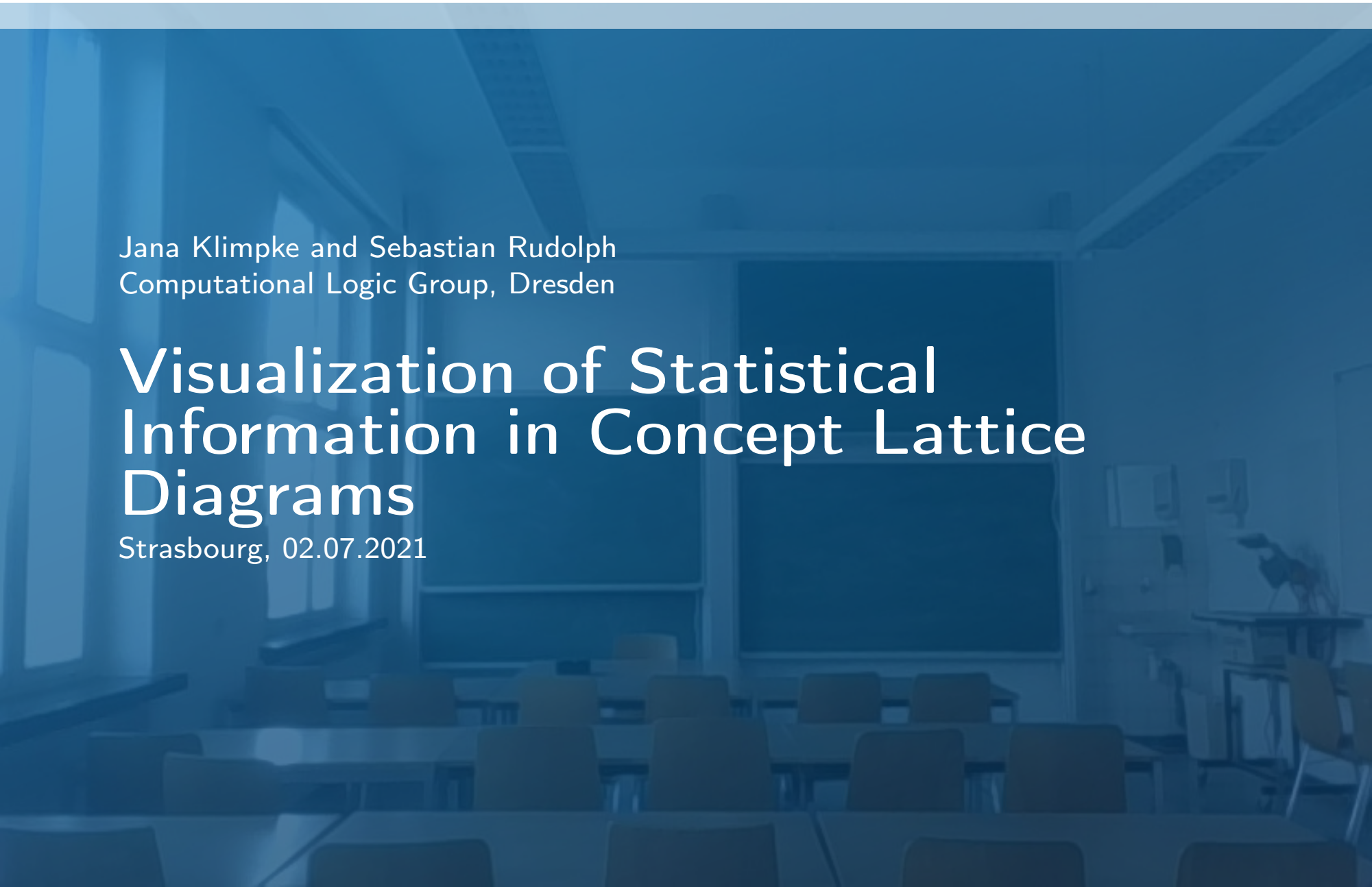


Jana Klimpke and Sebastian Rudolph
Computational Logic Group, Dresden

Visualization of Statistical Information in Concept Lattice Diagrams

Strasbourg, 02.07.2021

Table of Contents

Preliminaries - FCA

Formal Context (G, M, I)

- G - Set of *objects*
- M - Set of *attributes*
- $I \subseteq G \times M$ - Binary incidence relation between G and M

$$A' := \{m \in M \mid \forall\, g \in A : (g, m) \in I\} \tag{1}$$

$$B' := \{g \in G \mid \forall\, m \in B : (g, m) \in I\} \tag{2}$$

Formal Concept (A, B)

- Extent $A \subseteq G$
- Intent $B \subseteq M$

Order Relation $\leq$

- Used for hierarchically ordering of formal concepts.
- Using concepts (A_1, B_1) and (A_2, B_2): $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$.

Preliminaries - FCA

$G \setminus M$	a	b	c	d	e
T1		×		×	
T2		×			×
T3			×		
T4	×	×	×		
T5				×	
T6		×	×		
T7					×

Table 1: Cross table displaying the relation between objects and attributes of a given formal context.

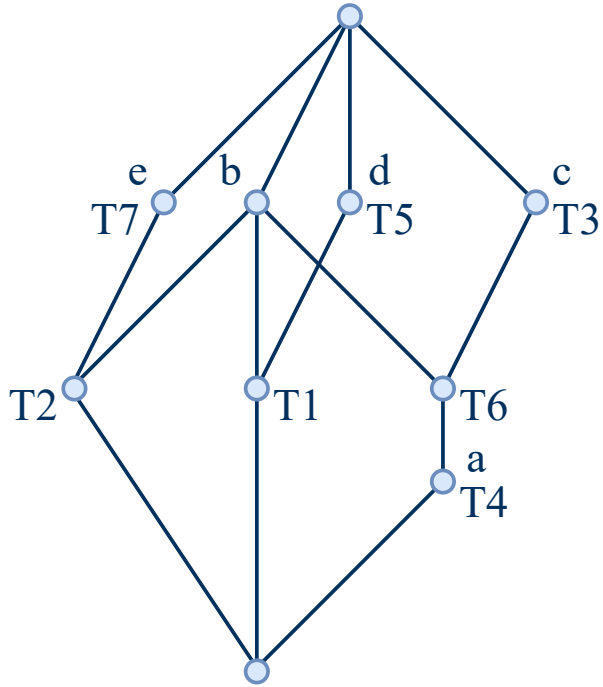


Figure 1: Example of an additive line diagram based on the context given in Table 1.

Implication

Rule of the form $X \to Y$, where X is the body and Y is the head of the rule. X and Y are sets of attributes. Given a formal context an implication $X \to Y$ is valid iff $Y \subseteq X''$.

Preliminaries - Association Analysis

Association

- $X \Rightarrow Y$
- X, Y: disjoint and real subsets of the set of all attributes M
- An object fulfills such a rule if it has all the attributes, which occur in X and Y.

Support

$$supp(X) := \frac{|\{g \in G \mid X \subseteq \{g\}'\}|}{|G|} = \frac{|X'|}{|G|} \tag{3}$$

$$supp(X \Rightarrow Y) := supp(X \cup Y) = \frac{|(X \cup Y)'|}{|G|} \tag{4}$$

Preliminaries - Association Analysis

Confidence

$$conf(X \Rightarrow Y) := \frac{supp(X \Rightarrow Y)}{supp(X)} \tag{5}$$

Lift

$$lift(X \Rightarrow Y) := \frac{conf(X \Rightarrow Y)}{supp(Y)} = \frac{supp(X \cup Y)}{supp(X) \cdot supp(Y)} \tag{6}$$

- $lift(X \Rightarrow Y) > 1$: X and Y positively correlated
- $lift(X \Rightarrow Y) = 1$: X and Y not correlated
- $lift(X \Rightarrow Y) < 1$: X and Y negatively correlated

Weighted Formal Contexts

$$mult : G \to \mathbb{N}^+ \tag{7}$$

$$wgt : \mathfrak{P}(G) \to \mathbb{N} \tag{8}$$

$$wgt(A) := \sum_{g \in A} mult(g) \tag{9}$$

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

Table 2: Example of a weighted context.

Linear Vertical Positioning

Goal

Read support, confidence and lift from the line diagram.

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

Table 2: Example of a weighted context.

Linear Vertical Positioning

Goal

Read support, confidence and lift from the line diagram.

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

Table 2: Example of a weighted context.

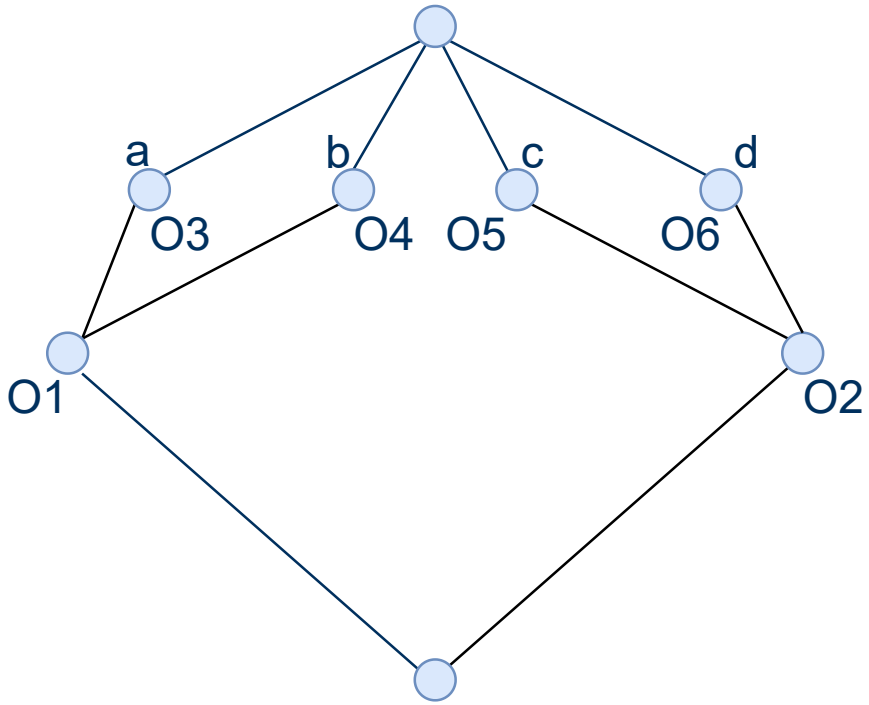


Figure 2: Additive line diagram.

Linear Vertical Positioning

Goal

Read support, confidence and lift from the line diagram.

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

Table 2: Example of a weighted context.

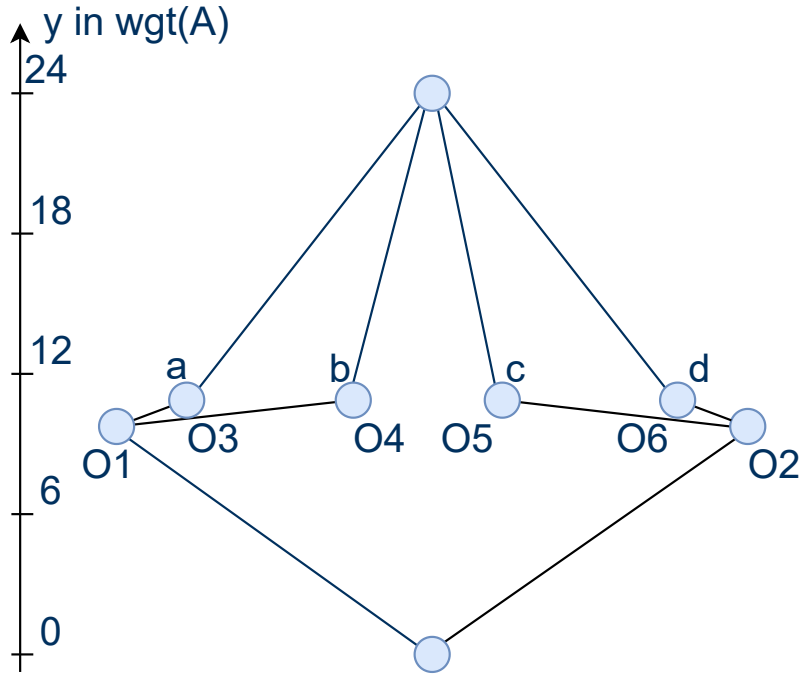


Figure 3: Y-weighted, additive line diagram.

Linear Vertical Positioning

Goal

Read support, confidence and lift from the line diagram.

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

Table 2: Example of a weighted context.

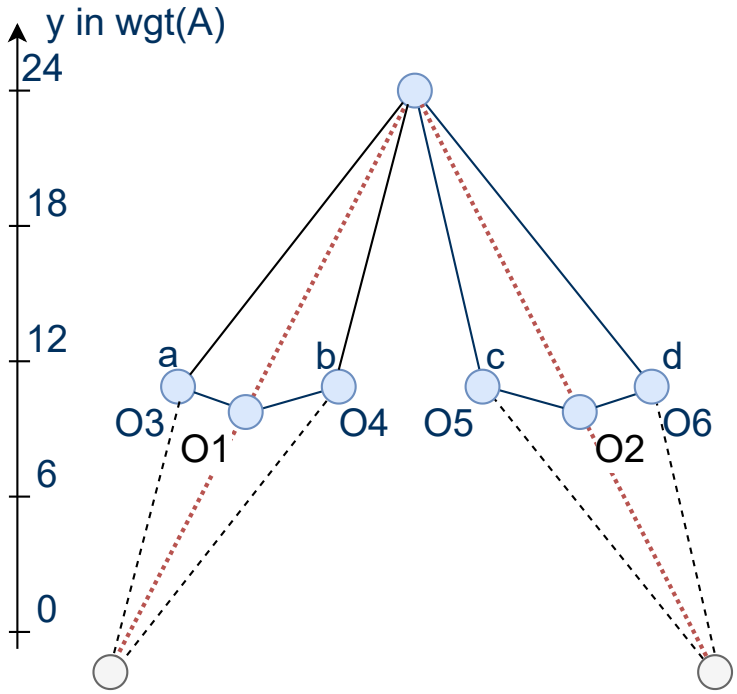


Figure 4: Weighted line diagram with auxiliary lines.

Linear Vertical Positioning

Goal

Read support, confidence and lift from the line diagram.

	mult	a	b	c	d
O1	10	×	×		
O2	10			×	×
O3	1	×			
O4	1		×		
O5	1			×	
O6	1				×

Table 2: Example of a weighted context.

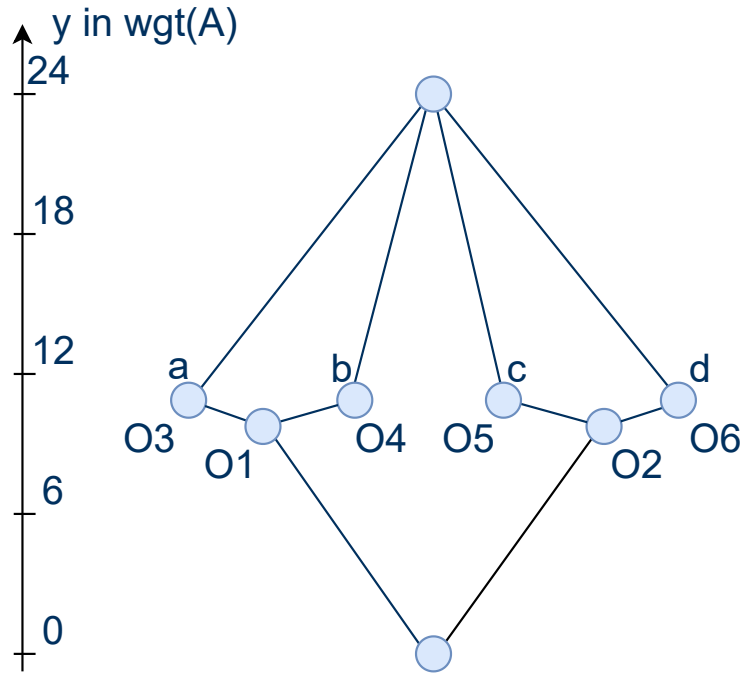


Figure 5: X-Y-weighted, additive line diagram.

Linear Vertical Positioning

Goal

Read support, confidence and lift from the line diagram.

Accomplished (so far)

Read support of a **formal concept** from the line diagram.

Linear Vertical Positioning

Support - Association

$$supp(X \Rightarrow Y) := supp(X \cup Y) = \frac{|(X \cup Y)'|}{|G|} \tag{4}$$

Support - Implication

$$supp(X \rightarrow Y) := \frac{supp(X' \cap Y')}{supp(G)} = \frac{supp(X \cup Y)}{supp(G)} \tag{7}$$

Logarithmized support

$$\log_a(supp(X)) = \log_a(supp(X \cup Y)) - \log_a(supp(G)) \tag{8}$$

Linear Vertical Positioning

	mult	male	female	<175	≥ 175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Table 3: Distribution by height and gender in Germany 2006.[1]

[1] https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/

Linear Vertical Positioning

	mult	male	female	<175	≥ 175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Table 3: Distribution by height and gender in Germany 2006.[1]

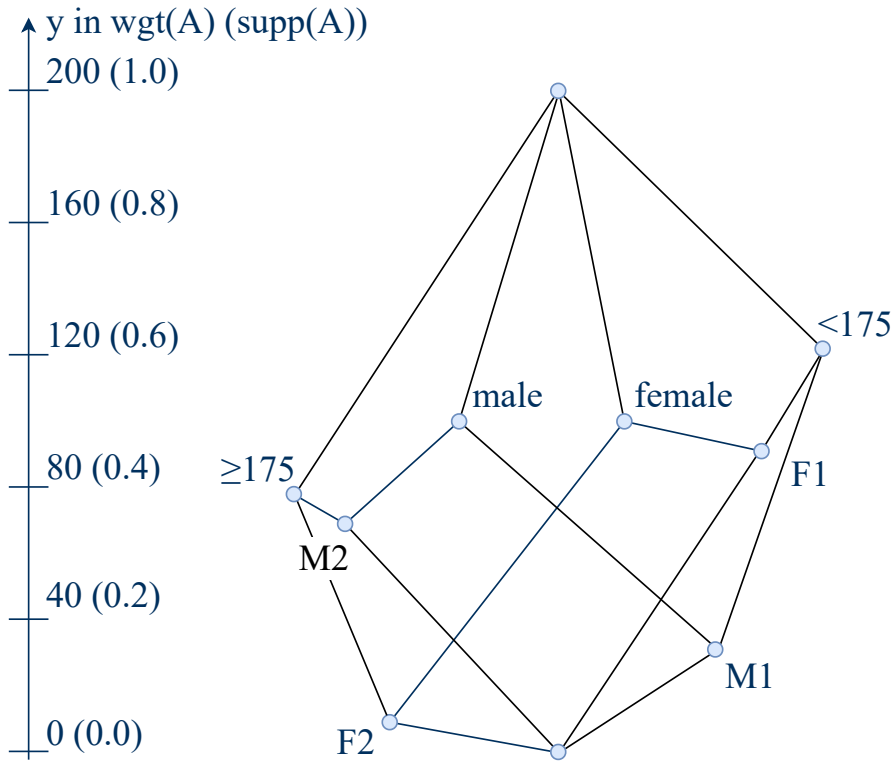


Figure 6: Gender and size distribution with representation of $wgt(A)$ and $supp(A)$.

[1] https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/

Linear Vertical Positioning

	mult	male	female	<175	≥ 175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Table 3: Distribution by height and gender in Germany 2006.[1]

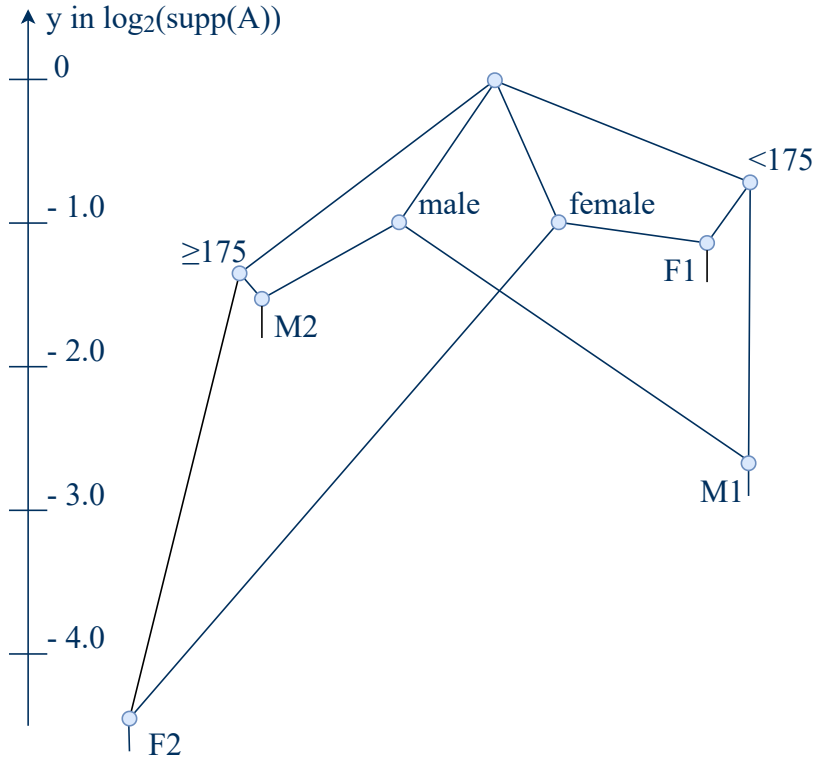


Figure 7: Gender and size distribution with representation of the logarithmized support of each node on the y-axis.

[1] https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/

Linear Vertical Positioning

	mult	male	female	<175	≥ 175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Table 3: Distribution by height and gender in Germany 2006.[1]

$$\{male\} \rightarrow \{\geq 175\}$$

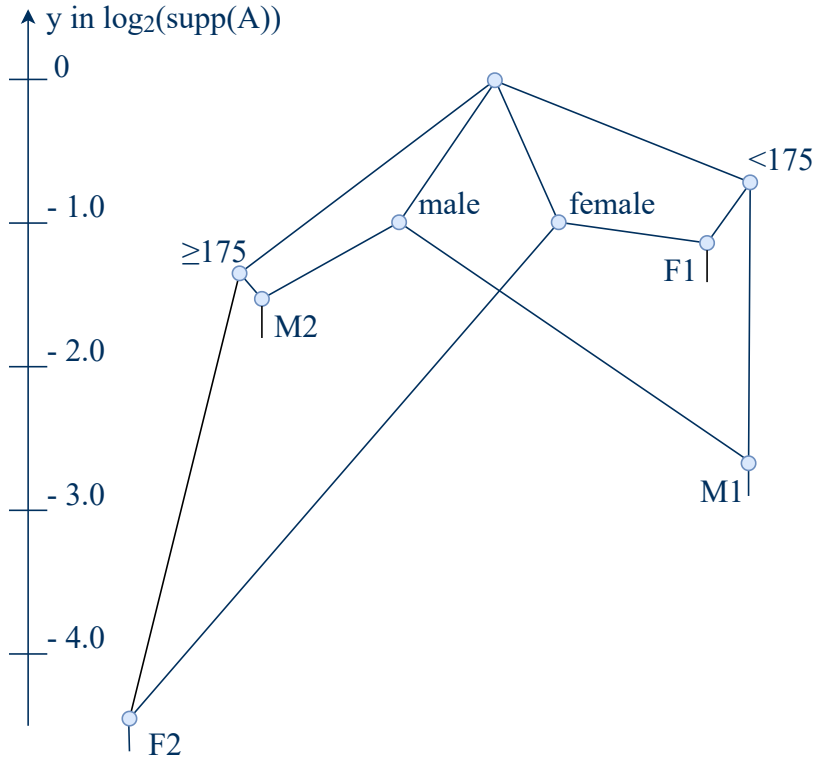


Figure 7: Gender and size distribution with representation of the logarithmized support of each node on the y-axis.

[1] https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/

Linear Vertical Positioning

	mult	male	female	<175	≥ 175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Table 3: Distribution by height and gender in Germany 2006.[1]

$$\{male\} \rightarrow \{\geq 175\}$$

$$\log_a(conf(X \rightarrow Y)) := \log_a(supp(X \rightarrow Y)) - \log_a(supp(X))$$

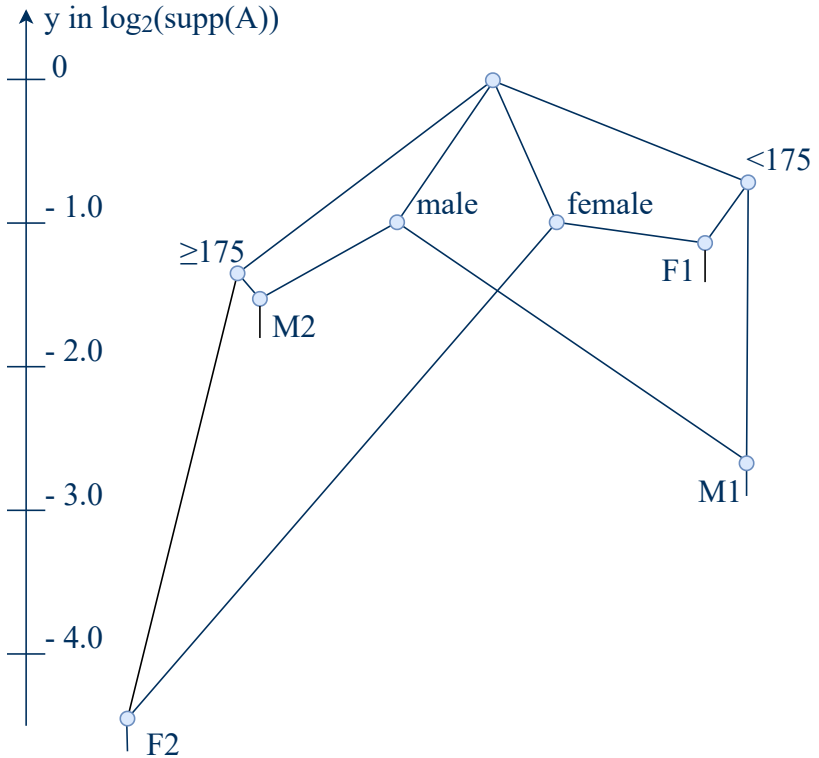


Figure 7: Gender and size distribution with representation of the logarithmized support of each node on the y-axis.

[1] https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/

Linear Vertical Positioning

	mult	male	female	<175	≥ 175
M1	31	×		×	
M2	69	×			×
F1	91		×	×	
F2	9		×		×

Table 3: Distribution by height and gender in Germany 2006.[1]

$$\{male\} \rightarrow \{\geq 175\}$$

$$\log_a(conf(X \rightarrow Y)) := \log_a(supp(X \rightarrow Y)) - \log_a(supp(X))$$

$$\log_a(lift(X \rightarrow Y)) := \log_a(conf(X \rightarrow Y)) - \log_a(supp(X))$$

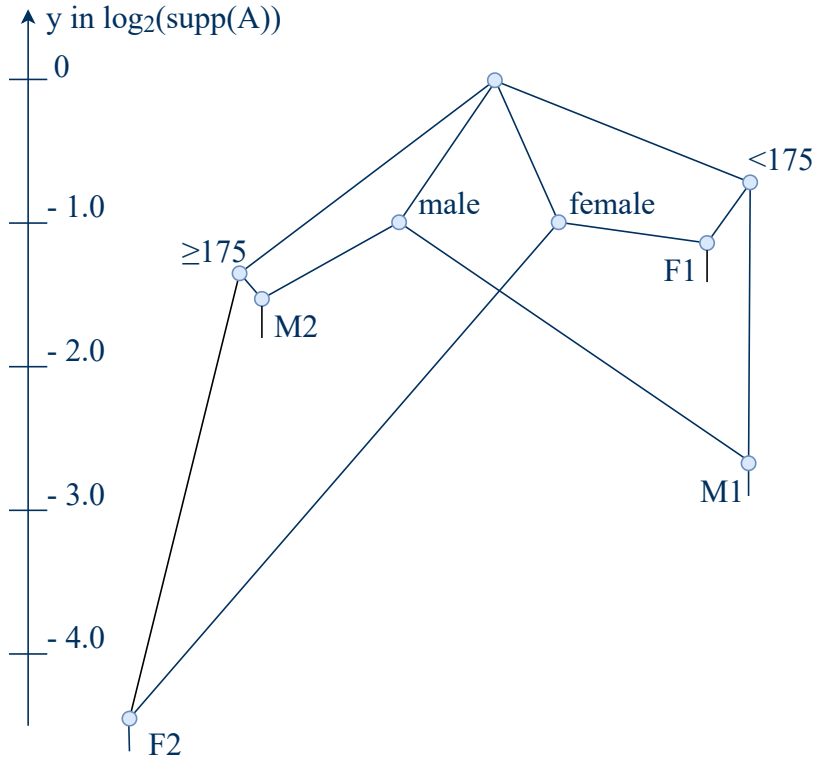


Figure 7: Gender and size distribution with representation of the logarithmized support of each node on the y-axis.

[1] https://de.statista.com/statistik/daten/studie/1825/umfrage/koerpergroesse-nach-geschlecht/

Reading Correlation - Numeric

$$lift(X \to Y) := \frac{conf(X \to Y)}{supp(Y)} \tag{9}$$

$$\frac{conf(X \to Y)}{supp(Y)} > 1.0 \tag{10}$$

$$\log_2\left(\frac{conf(X \to Y)}{supp(Y)}\right) > \log_2(1.0) \tag{11}$$

$$\log_2(conf(X \to Y)) - \log_2(supp(Y)) > 0.0 \tag{12}$$

$$\log_2(conf(X \to Y)) > \log_2(supp(Y)) \tag{13}$$

Reading Correlation - Numeric

Correlation between $\{male\}$ and $\{\geq 175\}$?

Compare $\log_2(conf(\{\geq 175\} \rightarrow \{male\}))$ and $\log_2(supp(\{male\}))$.

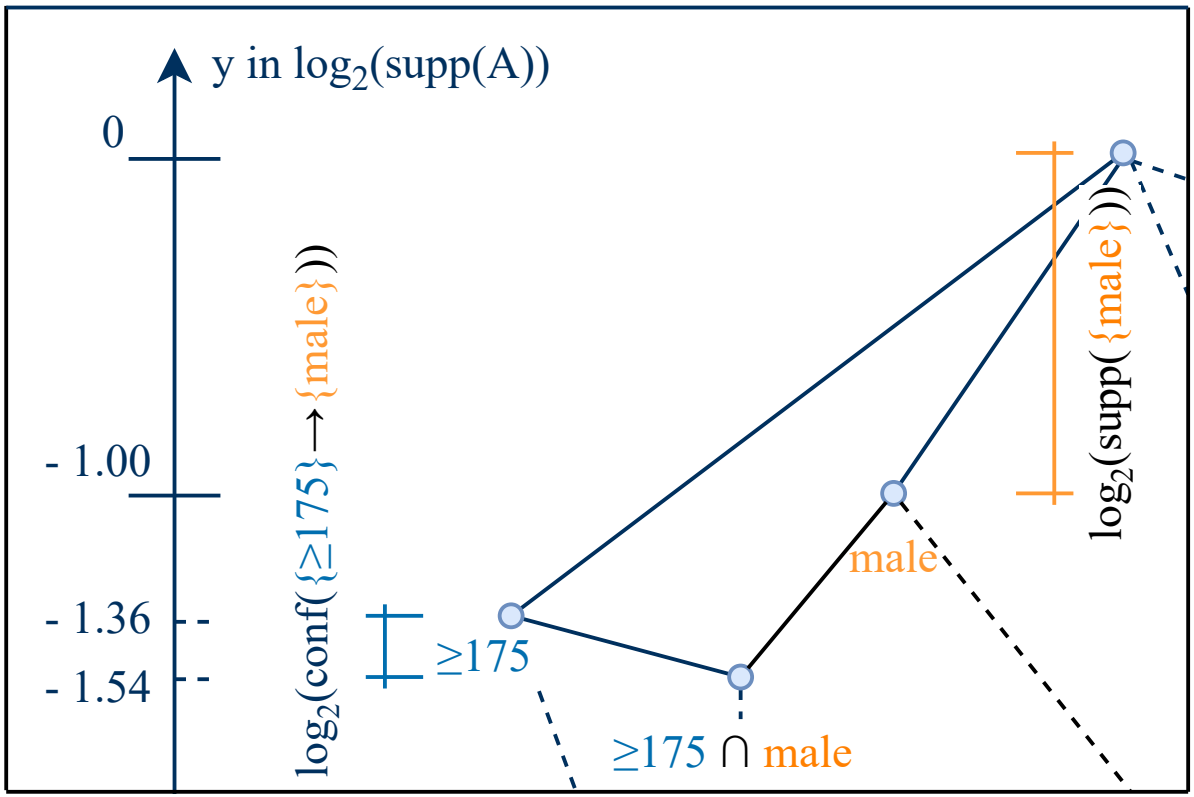


Figure 8: Estimating the dependence of X and Y by reading off confidence and support.

Reading Correlation - Parallelogram Method

Correlation between $\{male\}$ and $\{\geq 175\}$?

Check whether the formed quadrangle is well-formed, stretched or compressed.

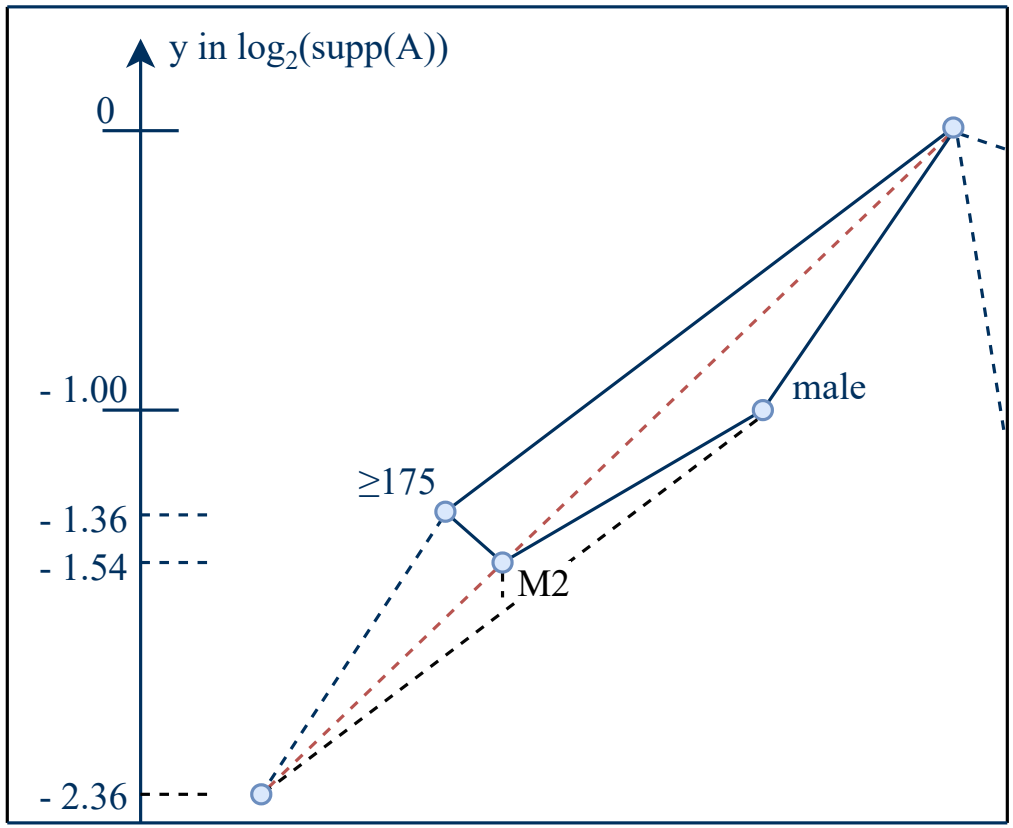


Figure 9: Illustrating the parallelogram method. The downward-pointing half of the quadrangle is compressed, indicating that $\geq$ 175 and *male* are positively correlated.

Reading Correlation - Parallelogram Method

Check whether the formed quadrangle is **well-formed**, stretched or compressed.

$$lift(X \to Y) := \frac{conf(X \to Y)}{supp(Y)}$$
$$:= \frac{supp(X \to Y)}{supp(Y) \cdot supp(X)}$$

	mult	a	b
T	20		
A	20	×	
B	20		×
AB	20	×	×

Table 4: Well-formed example.

$$lift(\{a\} \to \{b\}) = \frac{1/4}{1/2 \cdot 1/2} \quad (14)$$
$$= 1.0 \quad (15)$$

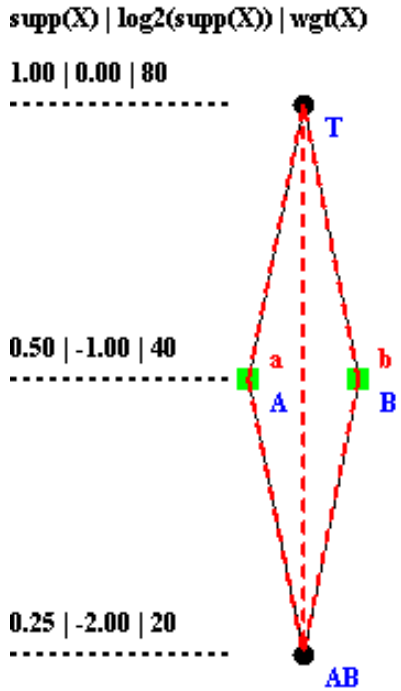


Figure 10: A, B independent.

Reading Correlation - Parallelogram Method

Check whether the formed quadrangle is well-formed, **stretched** or compressed.

$$lift(X \to Y) := \frac{conf(X \to Y)}{supp(Y)}$$
$$:= \frac{supp(X \to Y)}{supp(Y) \cdot supp(X)}$$

	mult	a	b
T	10		
A	30	×	
B	30		×
AB	10	×	×

Table 5: Stretched example.

$$lift(\{a\} \to \{b\}) = \frac{1/8}{1/2 \cdot 1/2} \quad (16)$$
$$= 1/2 < 1.0 \quad (17)$$

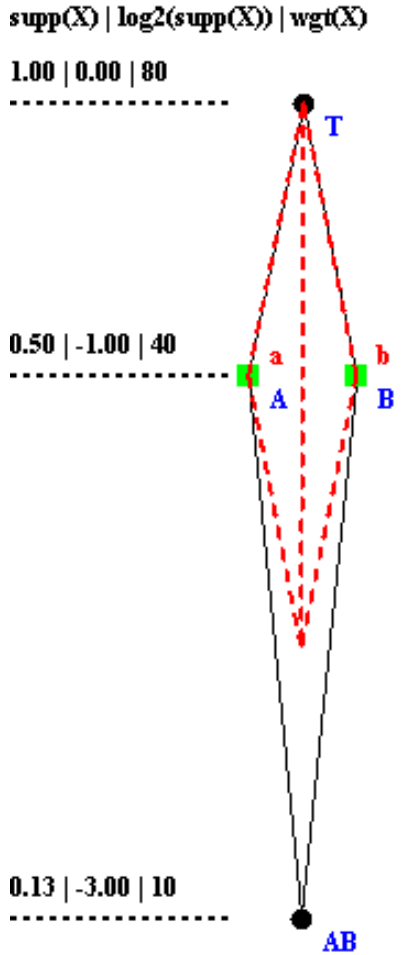


Figure 11: A, B neg. correlated.

Reading Correlation - Parallelogram Method

Check whether the formed quadrangle is well-formed, stretched or **compressed**.

$$lift(X \to Y) := \frac{conf(X \to Y)}{supp(Y)}$$
$$:= \frac{supp(X \to Y)}{supp(Y) \cdot supp(X)}$$

	mult	a	b
T	30		
A	10	×	
B	10		×
AB	30	×	×

Table 6: Compressed example.

$$lift(\{a\} \to \{b\}) = \frac{3/8}{1/2 \cdot 1/2} \quad (18)$$
$$= 3/2 > 1.0 \quad (19)$$

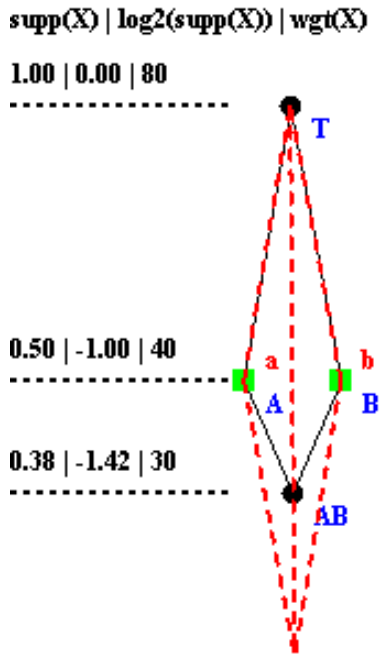


Figure 12: A, B pos. correlated.

Prototype

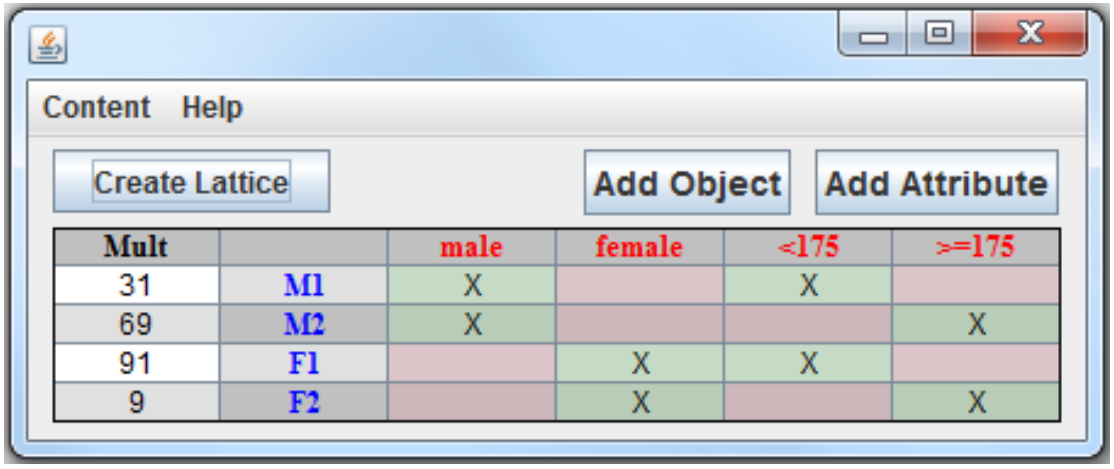


Figure 13: Display of a cross table in the program.

Prototype

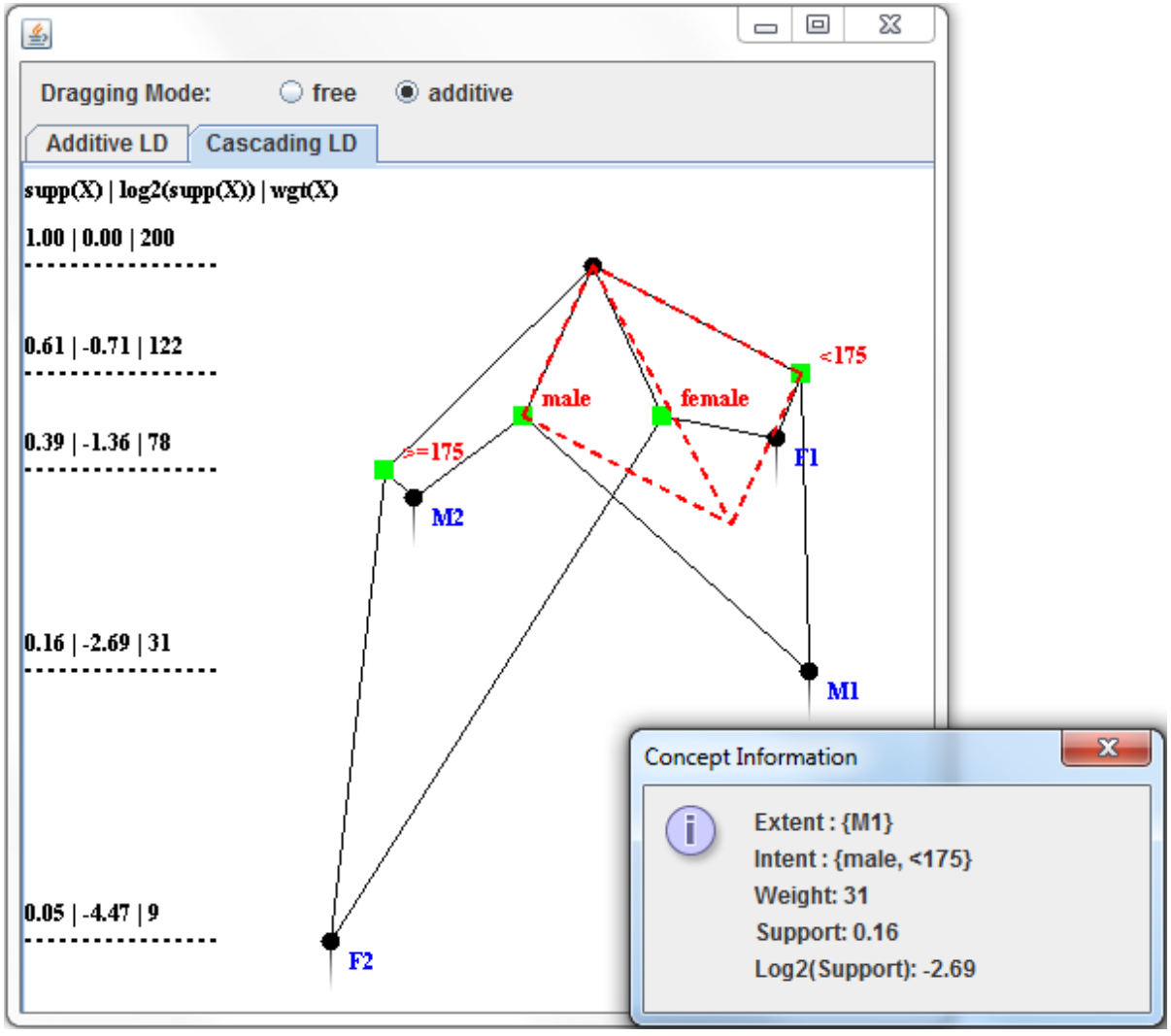
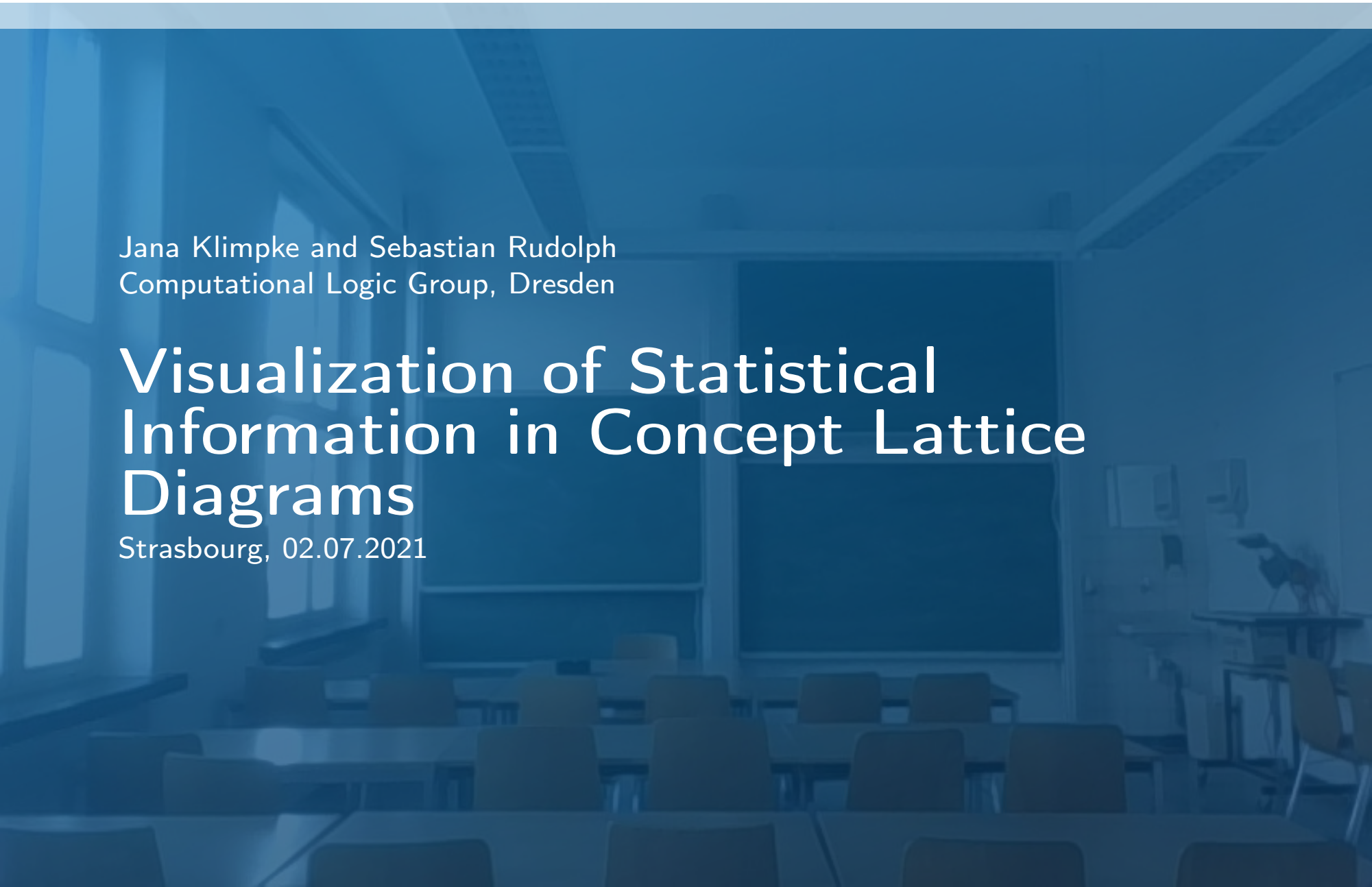


Figure 14: Display of the output of the program.

Summary

- Combined FCA with association analysis
- Using weighted contexts
- Visualized statistical relationships between attributes in diagrams:
  - Well-formed quadrangles == independent
  - Stretched quadrangles == neg. correlated
  - Compressed quadrangles == pos. correlated
- Introduced cascading additive line diagrams (see paper)
- Prototype in Java[1]

[1] `https://github.com/Klimpke/Cascading-Line-Diagrams-Visualizer`



Jana Klimpke and Sebastian Rudolph
Computational Logic Group, Dresden

Visualization of Statistical Information in Concept Lattice Diagrams

Strasbourg, 02.07.2021