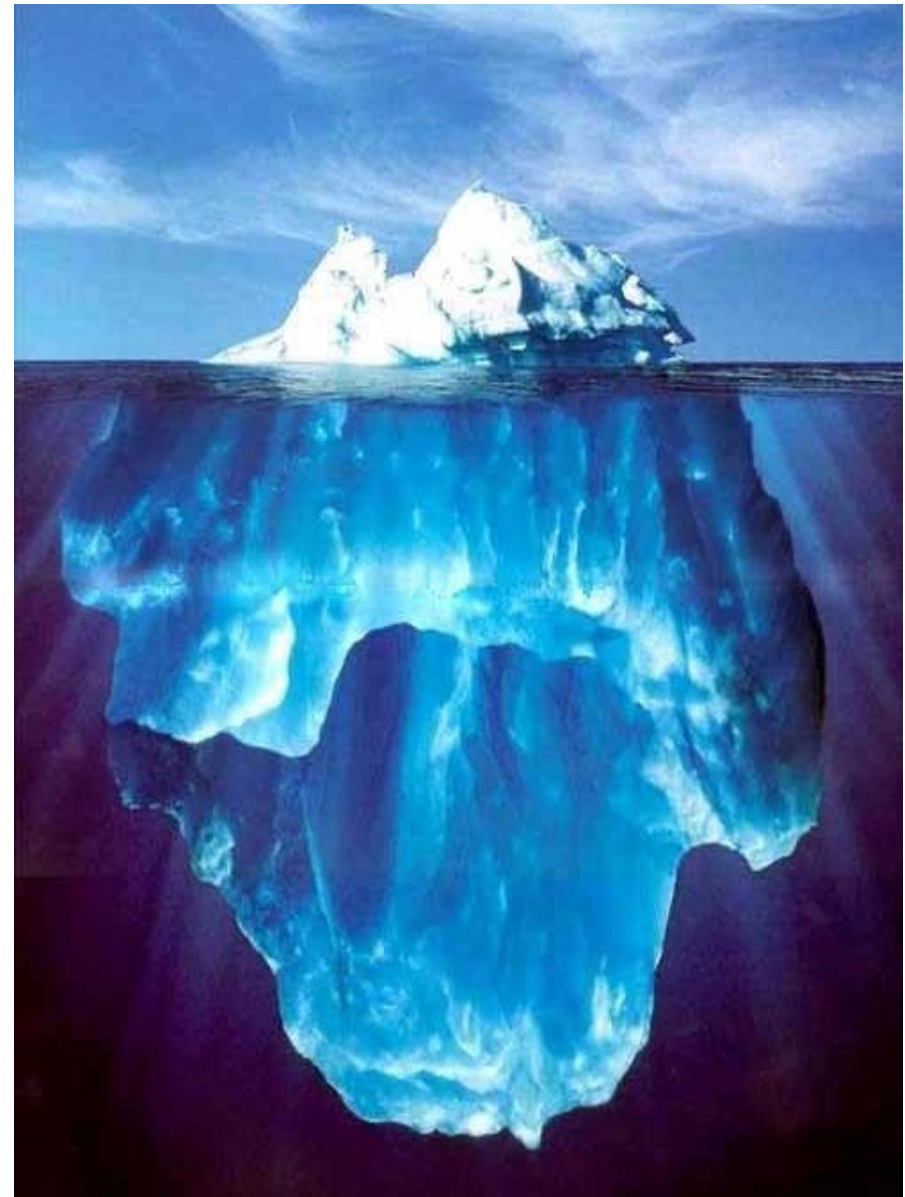


Formale Begriffsanalyse

Teil 2:

- Assoziationsregeln und FBA
- Titanic Algorithmus
- Begriffliches Skalieren
(= quantitative Assoziationsregeln)



Bases of Association Rules

{ veil color: white, gill spacing: close } → { gill attachment: free }

Support: 78,52 % Confidence: 99,6 %

Classical Data Mining Task: Find, for given minsupp, minconf $\in [0,1]$, all rules with support and confidence above these thresholds

Our task: Find a **basis** of rules, i.e., a minimal set of rules out of which all other rules can be derived.

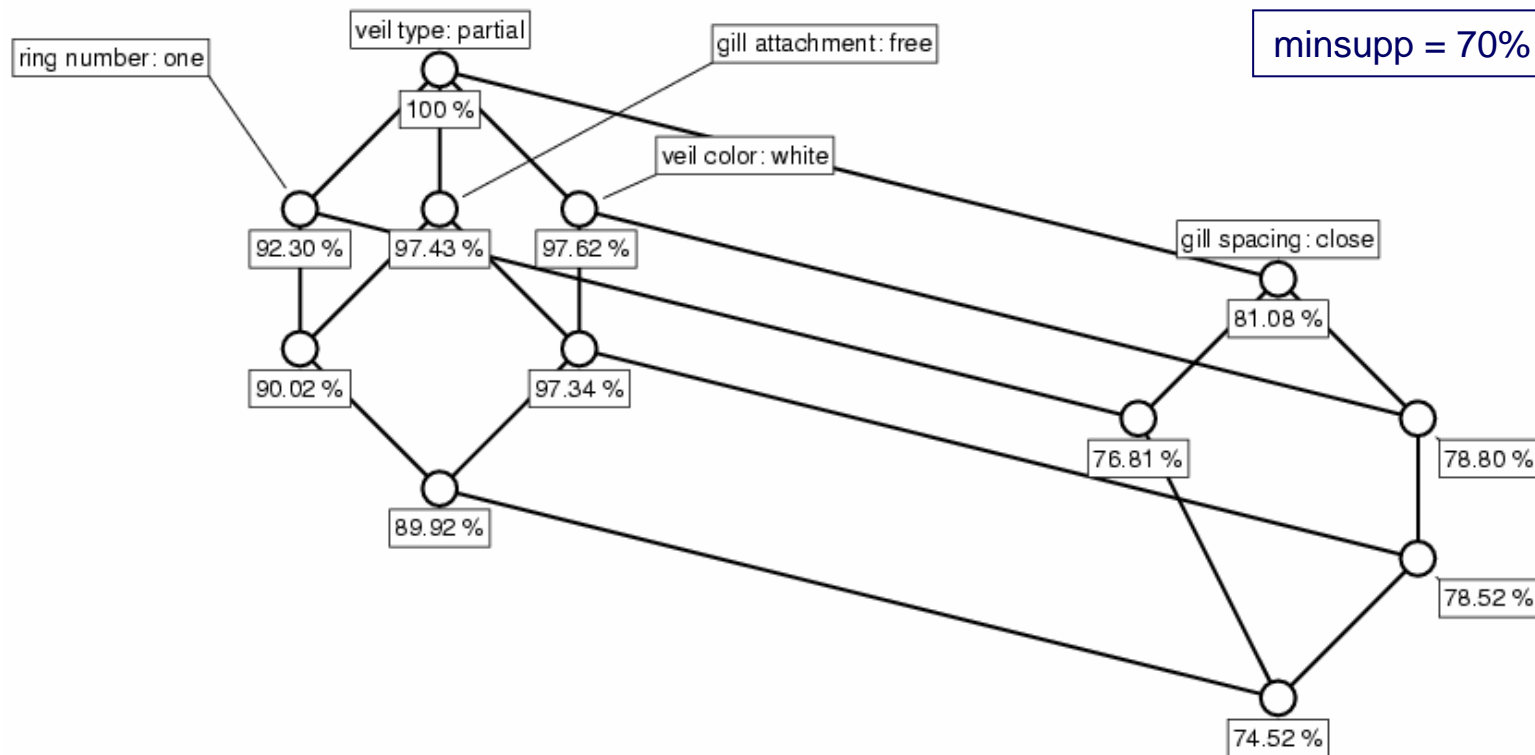
-
- From $B' = B''$ follows

$$\text{supp}(B) = \frac{|B'|}{|G|} = \frac{|B''|}{|G|} = \text{supp}(B')$$

Theorem: $X \rightarrow Y$ and $X'' \rightarrow Y''$ have the same support and the same confidence.

Hence for computing association rules, it is sufficient to compute the supports of all frequent sets with $B = B''$ (i.e., the intents of the iceberg concept lattice).

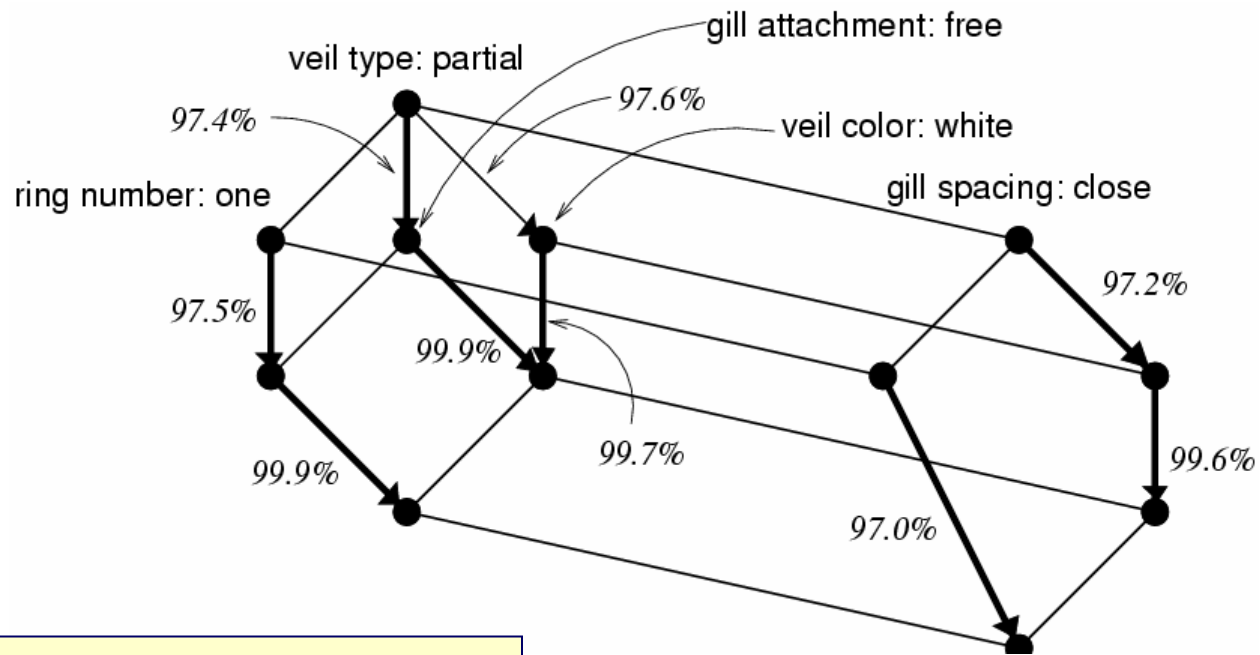
Advantage of the use of iceberg concept lattices (compared to frequent itemsets)



32 frequent itemsets are represented by 12 frequent concept intents

- more efficient computation (e.g. TITANIC)
- fewer rules (without information loss!)

Advantage of the use of iceberg concept lattices (compared to frequent itemsets)



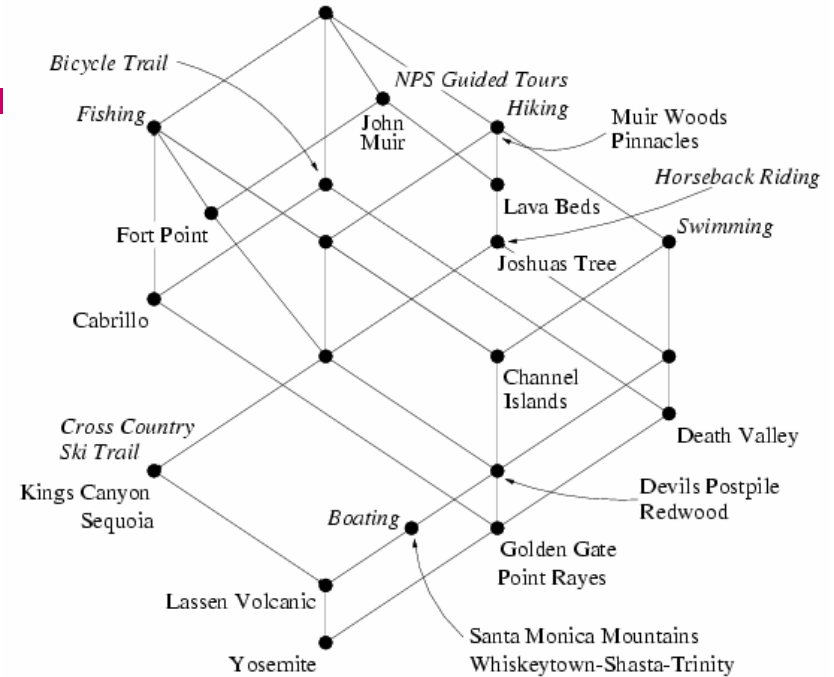
Association rules can be visualized in the iceberg concept lattice:

- **exact rules** ← conf = 100 %
- **approximate rules** ← conf < 100 %

Wdh.: Exakte Assoziationsregeln

In concept lattices, they can be directly read from the diagram:

- **Lemma:** An implication $X \rightarrow Y$ holds iff the largest concept which is below all concepts generated by the attributes in X is below all concepts generated by attributes in Y .

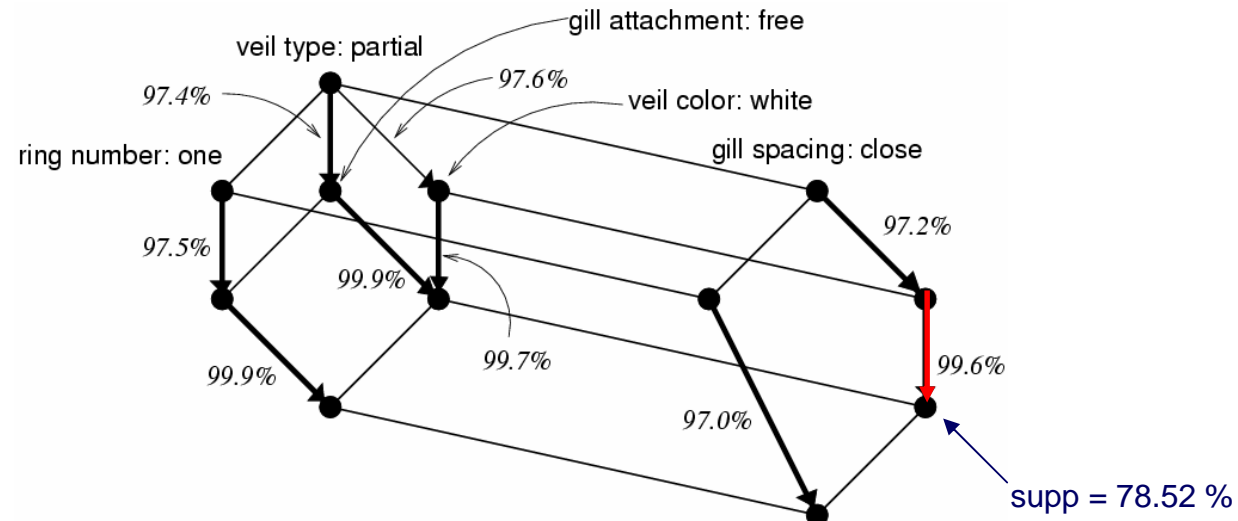


Examples:

- $\{ \text{Swimming} \} \rightarrow \{ \text{Hiking} \}$
(supp=10/19 \approx 52.6%, conf = 100%)
- $\{ \text{Boating} \} \rightarrow \{ \text{Swimming, Hiking, NPS Guided Tours, Fishing} \}$
(supp=4/19 \approx 21.0%, conf = 100%)
- $\{ \text{Bicycle Trail, NPS Guided Tours} \} \rightarrow \{ \text{Swimming, Hiking} \}$ (supp=4/19 \approx 21.0%, conf = 100%)

Approximate Association Rules

Def.: The **Luxenburger basis** consists of all valid association rules $X \rightarrow Y$ such that there are concepts (A_1, B_1) and (A_2, B_2) where (A_1, B_1) is a direct upper neighbor of (A_2, B_2) , $X = B_1$, and $X \cup Y = B_2$.

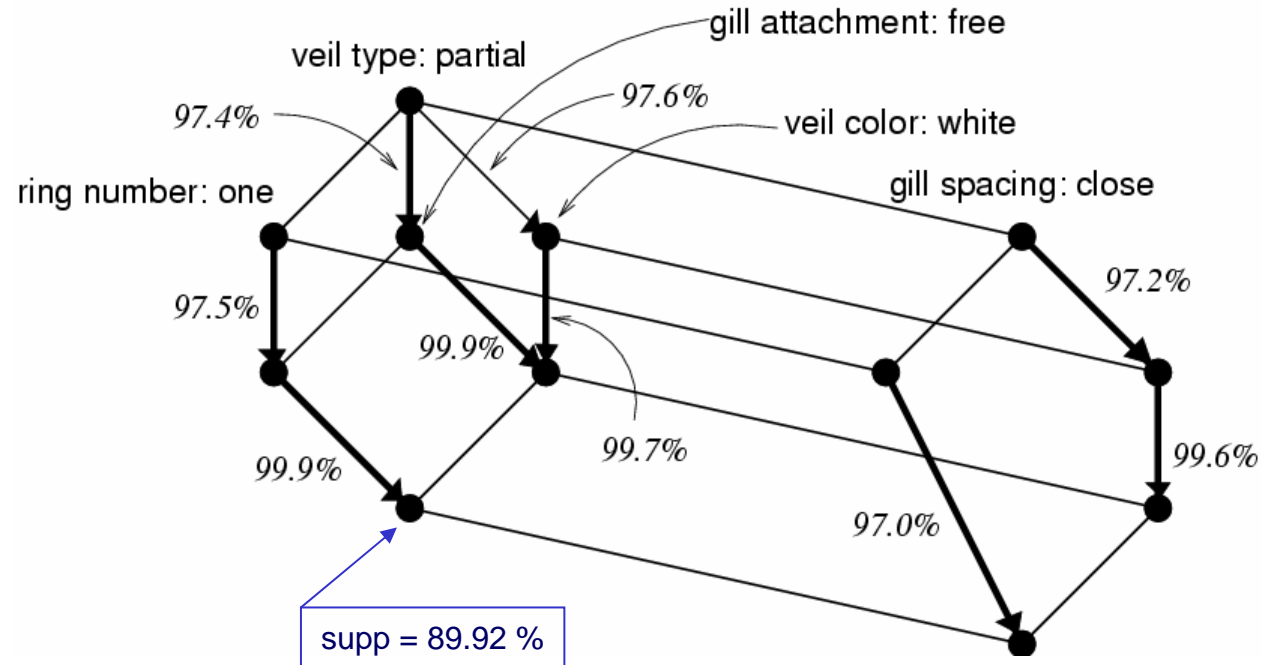


Each arrow indicates a rule of the basis, e.g. the rightmost arrow stands for $\{ \text{veil type: partial, gill spacing: close, veil color: white} \} \rightarrow \{ \text{gill attachment: free} \}$ (conf = 99.6 %, supp = 78.52 %)

Theorem: From the Luxenburger-Basis all approximate rules (incl. support und confidence) can be derived with the following rules:

- $\phi(X \rightarrow Y) = \phi(X \rightarrow Y \setminus Z)$, für $\phi \in \{ \text{conf}, \text{supp} \}$, $Z \subseteq X$
- $\phi(X'' \rightarrow Y'') = \phi(X \rightarrow Y)$
- $\text{conf}(X \rightarrow X) = 1$
- $\text{conf}(X \rightarrow Y) = p$, $\text{conf}(Y \rightarrow Z) = q \Rightarrow \text{conf}(X \rightarrow Z) = p \cdot q$
for all frequent concept intents $X \subset Y \subset Z$.
- $\text{supp}(X \rightarrow Z) = \text{supp}(Y \rightarrow Z)$, for all $X, Y \subseteq Z$.

The basis is minimal with this property.



Example:

{ ring number: one } → { veil color: white }

- has support 89.92 % (the support of the largest concept having both attributes in its intent)
- and confidence $97.5 \% \times 99.9 \% \approx 97.4 \%$.

Name	Number of objects	Average size of objects	Number of items
T10I4D100K	100,000	10	1,000
MUSHROOMS	8,416	23	127
C20D10K	10,000	20	386
C73D10K	10,000	73	2,177

Some experimental results

Dataset (Minsupp)	Exact rules	D.-G. basis	Minconf	Approximate rules	Luxenburger basis
T10I4D100K (0.5%)	0	0	90%	16,269	3,511
			70%	20,419	4,004
			50%	21,686	4,191
			30%	22,952	4,519
MUSHROOMS (30%)	7,476	69	90%	12,911	563
			70%	37,671	968
			50%	56,703	1,169
			30%	71,412	1,260
C20D10K (50%)	2,277	11	90%	36,012	1,379
			70%	89,601	1,948
			50%	116,791	1,948
			30%	116,791	1,948
C73D10K (90%)	52,035	15	95%	1,606,726	4,052
			90%	2,053,896	4,089
			85%	2,053,936	4,089
			80%	2,053,936	4,089

minimale Anzahl von exakten Regeln, aus denen alle abgeleitet werden können.

TITANIC

Only concepts with a support above a threshold $\text{minsupp} \in [0,1]$.

TITANIC computes the set of all (frequent) concept intents using the *support function*:

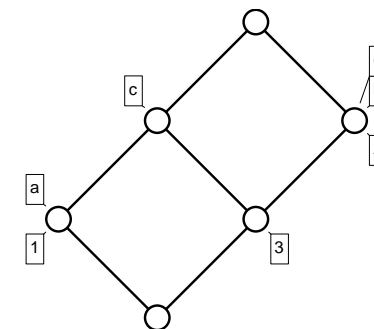
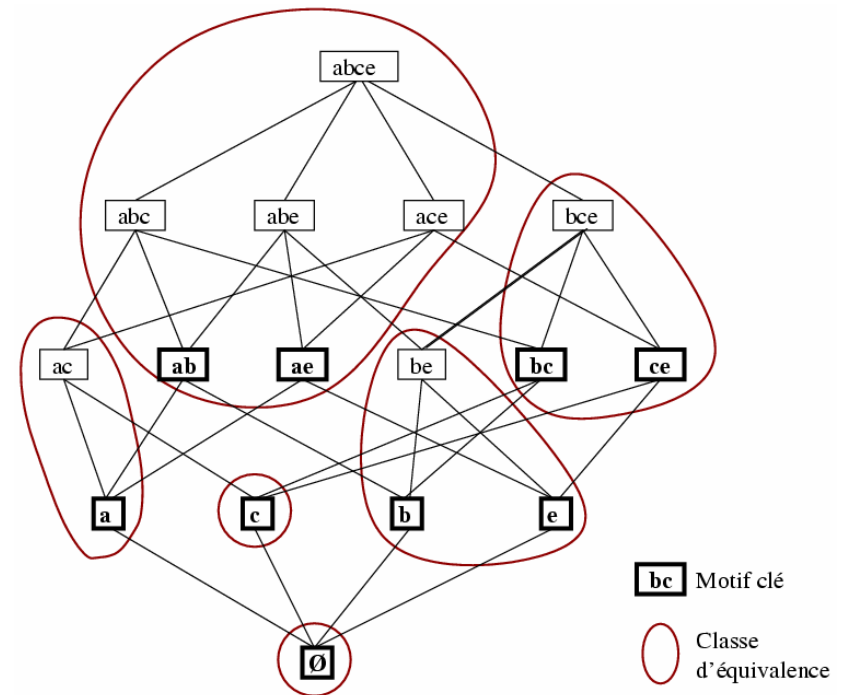
Def.: The **support** of an attribute set (itemset) $X \subseteq M$ is given by

$$\text{supp}(X) = \frac{|X|}{|G|}$$

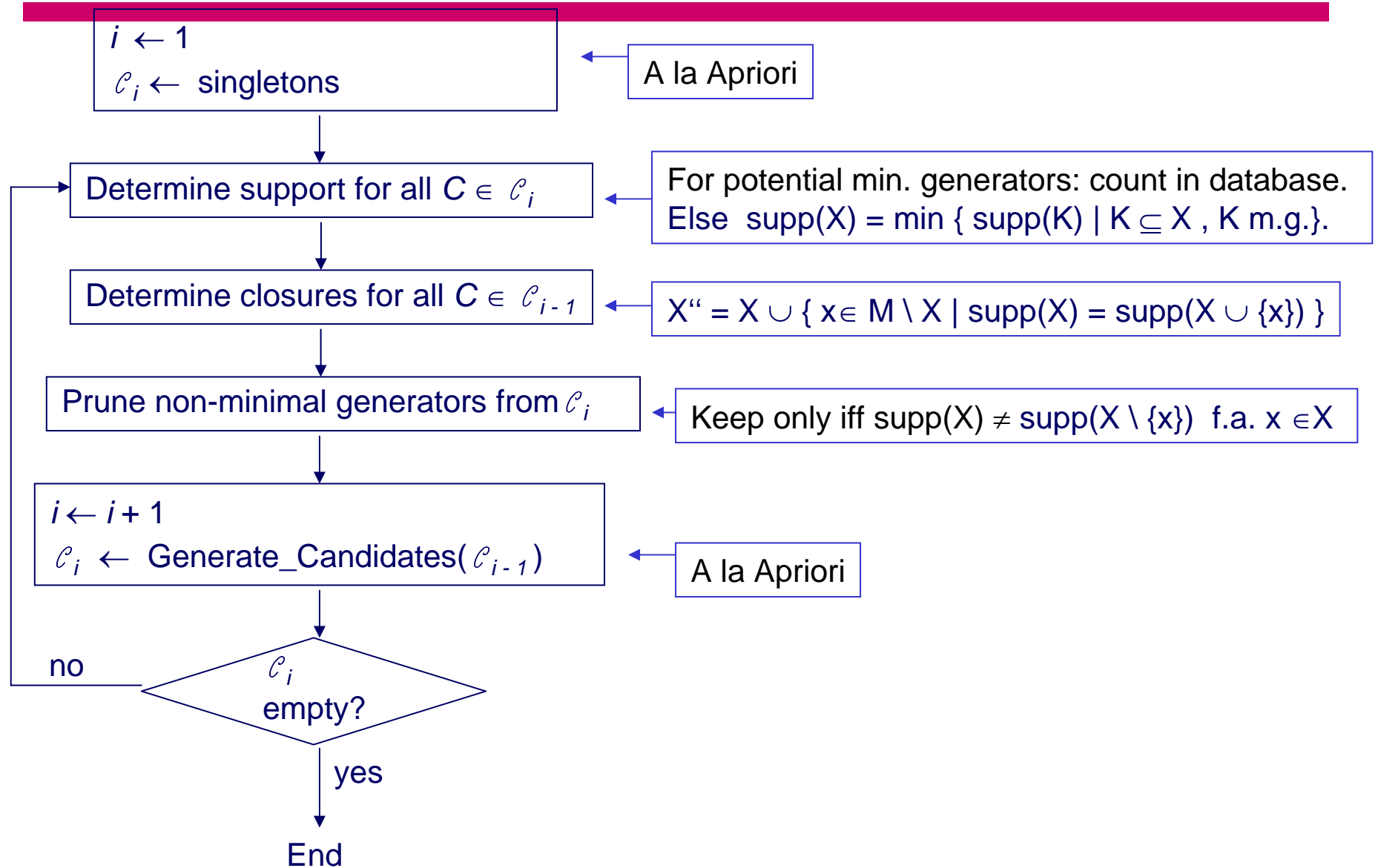
TITANIC

	a	b	c	e
1	×		×	
2		×		×
3		×	×	×

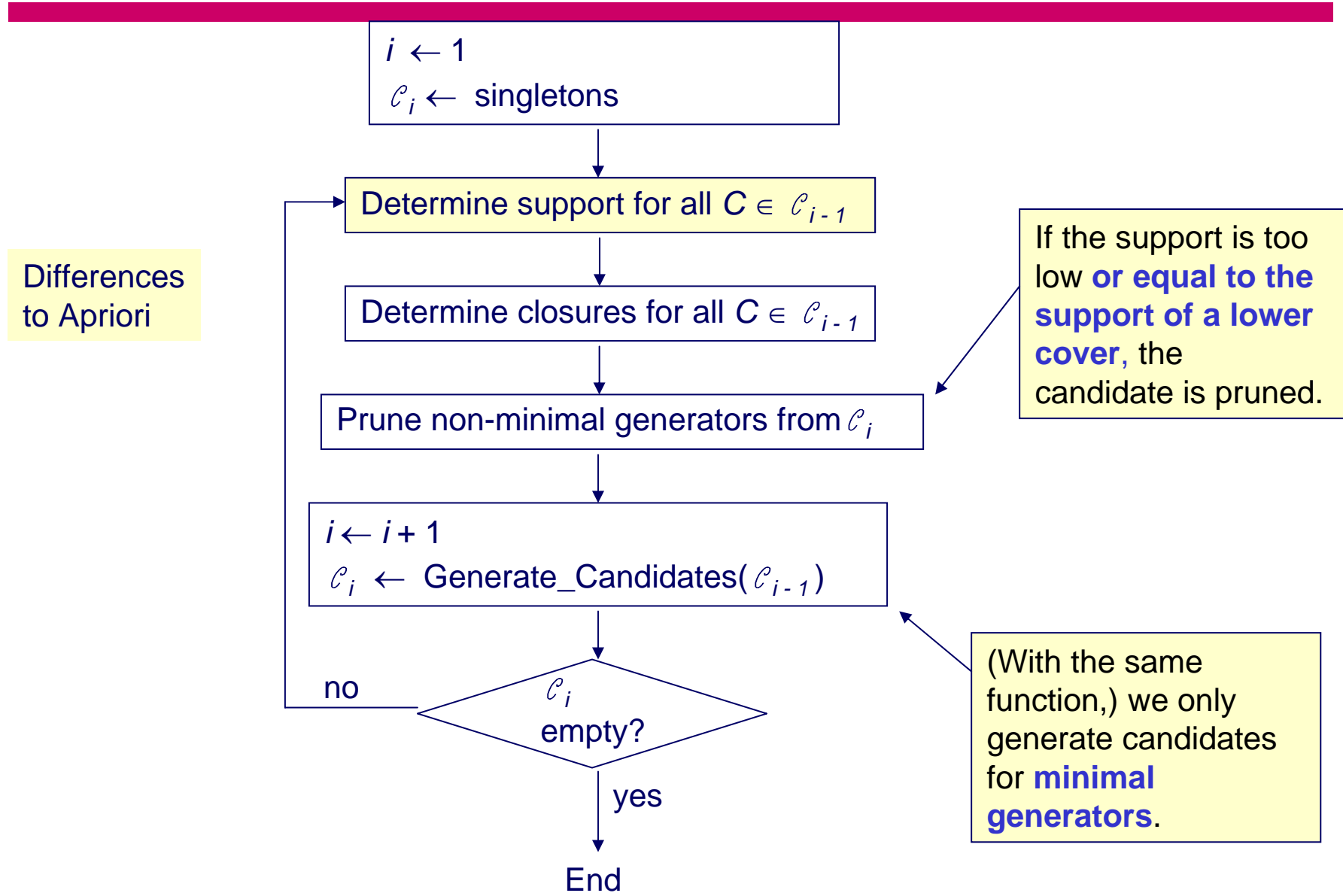
- Alle Itemsets in einer Äquivalenzklasse haben denselben Support (da sie in denselben Transaktionen enthalten sind).
- Für die Berechnung fassen wir nur die kleinsten Elemente einer Klasse an, die sogen. minimalen Erzeuger (die dicken schwarzen Kästen). Sie haben die selbe Monotonie-Eigenschaft wie häufige Itemsets.
- Für die Regeln verwenden wir nur die Begriffsinhalte (die größten Elemente jeder Äquivalenzklasse, die wir aus einem minimalen Erzeuger X durch $X'' = X \cup \{x \in M \setminus X \mid \text{supp}(X) = \text{supp}(X \cup \{x\})\}$ erhalten).



TITANIC



TITANIC



Mehrwertige Kontexte und Skalierung

Dieser Abschnitt ersetzt die quantitativen Assoziationsregeln.
Skalierung kann auch als eine Methode der Datenvorverarbeitung
verstanden werden.

Def.: Ein mehrwertiger Kontext (G, M, W, I) besteht aus den Mengen G , M und W und einer dreistelligen Relation I zwischen G , M und W (d.h. $I \subseteq G \times M \times W$), wobei gilt:

aus $(g, m, w) \in I$ und $(g, m, v) \in I$ folgt stets $w = v$.

Die Elemente von G nennen wir die **Gegenstände**, die von M die **(mehrwertigen) Merkmale** und die von W die **Merkmalsausprägungen** oder **Werte**. $(g, m, w) \in I$ lesen wir als „ das Merkmal m hat beim Gegenstand g den Wert w “.

Bsp.: Der mehrwertige Kontext enthält einen Vergleich der verschiedenen Möglichkeiten, Motor und Antrieb eines PKW anzuordnen.



Standardantrieb



Frontantrieb



Heckantrieb



Mittelmotor mit
Heckantrieb



Allradantrieb

	Au	Ab	F	E	R	B	W
Standard	schlecht	gut	gut	untersteuernd	gut	mittel	sehr gut
Front	gut	schlecht	sehr gut	untersteuernd	sehr gut	sehr gering	gut
Heck	sehr gut	sehr gut	sehr schlecht	übersteuernd	schlecht	gering	sehr schlecht
Mittel	sehr gut	sehr gut	gut	neutral	sehr schlecht	gering	sehr schlecht
Allrad	sehr gut	sehr gut	gut	unterst./neutral	gut	hoch	schlecht

Au := Antriebswirkung unbeladen; Ab := Antriebswirkung beladen; F := Fahrstabilität; E := Eigenlenkverhalten; R := Raumausnutzung; B := Bauaufwand; W := Wartungsfreundlichkeit.

[Antriebskonzept für Personenkraftwagen. Quelle: Schlag nach! 100 000 Tatsachen aus allen Wissenschaftsgebieten. BI-Verlag Mannheim, 1982]

Begriffliche Skalierung

= Umwandlung eines mehrwertigen Kontexts in einen einwertigen.

Zur Skalierung wird zunächst jedes Merkmal eines mehrwertigen Kontextes durch einen Kontext interpretiert, dieser wird *begriffliche Skala* genannt:

Def.: Eine **Skala** zum Merkmal m eines mehrwertigen Kontextes ist ein (einwertiger) Kontext $S_m := (G_m, M_m, I_m)$ mit $m(G) \subseteq G_m$. Die Gegenstände einer Skala nennen wir **Skalenwerte**, die Merkmale **Skalenmerkmale**.

Bem.: Jeder Kontext kann als Skala verwendet werden. Einige besonders einfache Kontexte werden immer wieder als Skalen benutzt.

Bsp.: Der folgende einwertige Kontext ergibt sich als der abgeleitete Kontext zu dem oben angegebenen mehrwertigen Kontext, wenn folgende Skalen benutzt werden:

		++	+	-
$S_{Au} := S_{Ab} :=$	++	x	x	
	+		x	
	-			x

		++	+	--
$S_F :=$	++	x	x	
	+		x	
	--			x

	u	ü	n	u/n
$S_E :=$	u	x		
	ü		x	
	n			x
	u/n			x

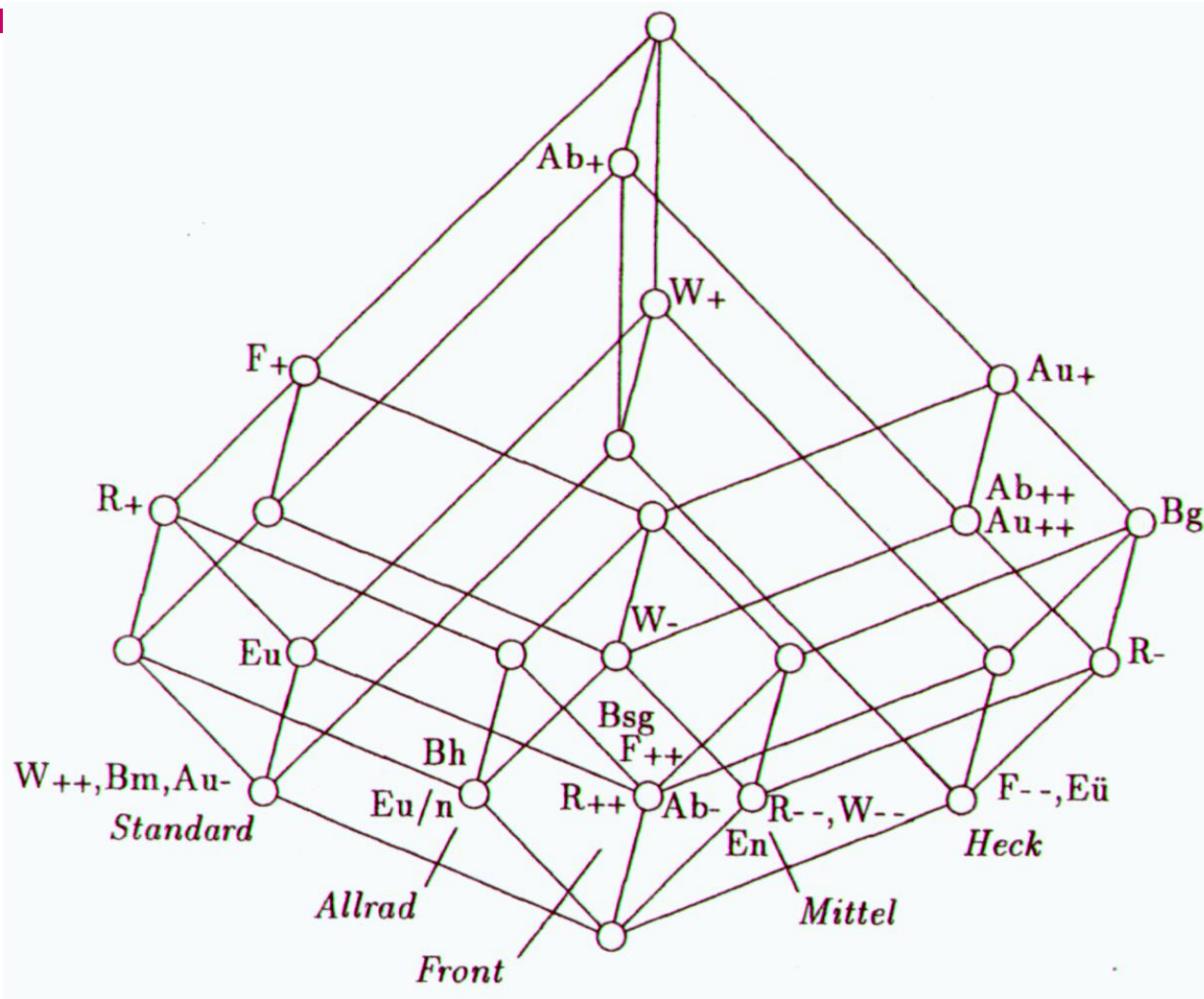
		++	+	-	--
$S_R := S_W :=$	++	x	x		
	+		x		
	-			x	
	--			x	x

		sg	g	m	h
$S_B :=$	sg	x	x		
	g		x		
	m			x	
	h				x

	Au			Ab			F			E				R				B				W			
	++	+	-	++	+	-	++	+	--	u	ü	n	u/n	++	+	-	--	sg	g	m	h	++	+	-	--
Standard			×		×			×		×					×					×		×	×		
Front		×				×	×	×		×				×	×			×	×				×		
Heck	×	×		×	×				×		×					×			×				×		
Mittel	×	×		×	×			×				×			×	×			×					×	×
Allrad	×	×		×	×			×					×		×						×			×	

Au := Antriebswirkung unbeladen; Ab := Antriebswirkung beladen; F := Fahrstabilität; E := Eigenlenkverhalten;
R := Raumausnutzung; B := Bauaufwand; W := Wartungsfreundlichkeit.
++ := sehr gut; + := gut; - := schlecht; -- := sehr schlecht; u := untersteuernd; Ü := übersteuernd; n := neutral;
sg := sehr gering; g := gering; m := mittel; h := hoch.

Hätten wir für die Merkmale *Au*, *Ab* und *F* ebenfalls die Skalen S_R verwendet, wäre der abgeleitete Kontext nur unwesentlich anders ausgefallen.



Begriffsverband zum Kontext der Antriebskonzepte

Bei der **begrifflichen Skalierung** erhält man aus dem mehrwertigen Kontext (G, M, W, I) und den Skalenkontexten $S_m, m \in M$, den abgeleiteten einwertigen Kontext wie folgt:

Die Gegenstandsmenge G bleibt unverändert, jedes mehrwertige Merkmal m wird ersetzt durch die Skalenmerkmale der Skala S_m . Denkt man sich einen mehrwertigen Kontext durch eine Tabelle dargestellt, so kann man sich die schlichte Skalierung folgendermaßen veranschaulichen: Jede Merkmalsausprägung $m(g)$ wird ersetzt durch die zu $m(g)$ gehörige Zeile des Skalenkontextes S_m .

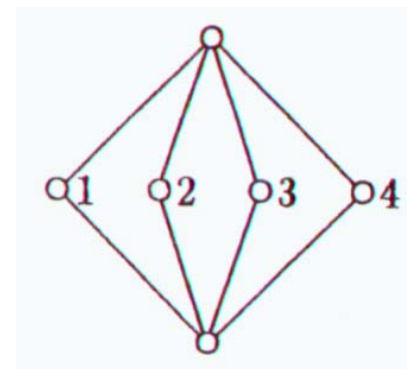
Def.: Elementarskalen

Nominalskalen $N_n := (n, n, =)$

Nominalskalen werden zur Skalierung von Merkmalen verwendet, deren Ausprägung sich gegenseitig ausschließen. Es ergibt sich eine Partition der Gegenstände in Begriffsumfänge.

	1	2	3	4
1	x			
2		x		
3			x	
4				x

Die Nominalskala N_4

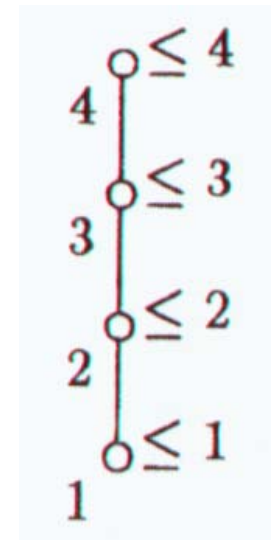


Ordinalskalen $O_n := (n, n, \leq)$

Ordinalskalen skalieren mehrwertige Merkmale, deren Ausprägungen geordnet sind und bei denen jede Merkmalsausprägung die jeweils schwächeren impliziert. Hat ein Merkmal z.B. die Ausprägungen $\{ \text{laut}, \text{sehr laut}, \text{extrem laut} \}$, so liegt es nahe, es ordinal zu skalieren. Die Merkmalsausprägungen bewirken dann eine Kette von Begriffsumfängen, die als *Rangordnung* gedeutet werden kann.

	1	2	3	4
1	x	x	x	x
2		x	x	x
3			x	x
4				x

Die Ordinalskala O_4



Interordinalskalen

$$I_n := (n, n, \leq) \mid (n, n, \geq)$$

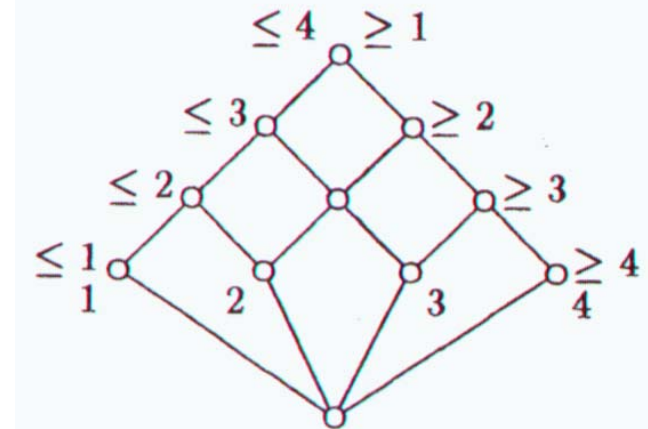
Die Begriffsumfänge der Interordinalskala sind genau die Intervalle von Ausprägungen.

Anwendungen:

- numerische Merkmale
- Gegensatzpaare (*aktiv-passiv, redselig-wortkarg* usw.) in Fragebögen mit Zwischenwerten.

$\Pi_4 =$

	≤ 1	≤ 2	≤ 3	≤ 4	≥ 1	≥ 2	≥ 3	≥ 4
1	x	x	x	x	x			
2		x	x	x	x	x		
3			x	x	x	x	x	
4				x	x	x	x	x



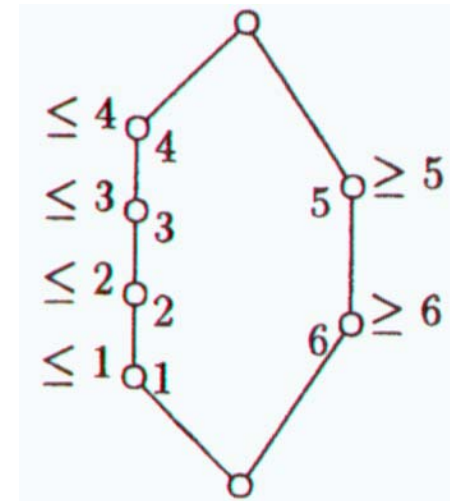
Bipolare Merkmale können aber oft auch treffend biordinal skaliert werden:

Biordinalskalen $M_{n,m} := (n, n, \leq) \cup (m, m, \geq)$

Im Sprachgebrauch verwenden wir Gegensatzpaare häufig nicht am Sinne der Interordinalskala, sondern einfacher: jedem Gegenstand wird einer der beiden Pole zugeordnet, wobei noch abgestuft werden kann. Die Ausprägungen $\{ \text{sehr leise, leise, laut, sehr laut} \}$ legen z.B. eine solche Skalierung nahe: laut und leise schließen sich aus, sehr laut impliziert laut, sehr leise impliziert leise. Eine entsprechende *Partition mit Rangordnung* hat man in den Namen der Schulnoten: Eine sehr gute Leistung ist natürlich auch gut, befriedigend und ausreichend, aber nicht mangelhaft oder ungenügend.

$M_{4,2} =$

	≤ 1	≤ 2	≤ 3	≤ 4	≥ 5	≥ 6
1	×	×	×	×		
2		×	×	×		
3			×	×		
4				×		
5					×	
6					×	×



Die **dichotome Skala** $D := (\{0, 1\}, \{0,1\}, =)$
wird häufig benutzt, um Merkmale mit { ja, nein } -Ausprägungen zu skalieren.

	0	1
0	x	
1		x

Bem.:

- Auf dem skalierten Kontext können nun natürlich wieder Assoziationsregeln berechnet werden.
- Wurden numerische Merkmale interordinal skaliert, so ergeben sich daraus sogen. ‚quantitative Assoziationsregeln‘.