IV) Generating the Cayley-tables themselves

We sketch how all Cayley-tables of semigroups on a given set S can be generated; mutatis mutandis everything generalizes to other types of universal algebras.

If i, j, k, α, β are elements of a semigroup $(S,*)$, then it follows from

$$\overbrace{i * j}^{=\alpha} * k \quad \text{that} \quad \alpha * k = i * \beta .$$
$$i * \underbrace{j * k}_{=\beta}$$

Slightly rephrasing:

$$\{ i*j=\alpha,\ j*k=\beta,\ \alpha*k=\gamma \} \longrightarrow \{ i*\beta=\gamma \}$$

Or better still:

$$\{ (i,j,\alpha), (j,k,\beta), (\alpha,k,\gamma) \} \longrightarrow \{ (i,\beta,\gamma) \}$$

In words: In the Cayley-table of a semigroup any three "strategically placed" entries of value α, β, γ imply that a certain fourth entry has value γ.

(24

Specifically, put $W := S \times S \times S$ and let Σ consist of the implications

$$\{(i,j,\alpha),(j,k,\beta),(\alpha,k,\gamma)\} \to \{(i,\beta,\gamma)\}$$

$$\{(i,j,\alpha),(j,k,\beta),(i,\beta,\gamma)\} \to \{(\alpha,k,\gamma)\}$$

where $i,j,k,\alpha,\beta,\gamma$ range freely over S.

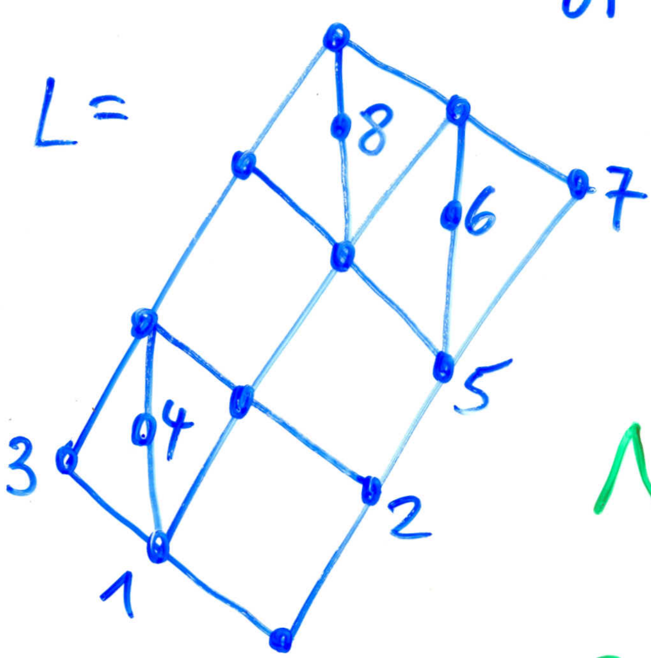
Then Σ is a (highly symmetric) implicational base for the closure system of all associative Cayley-tables on the set W.

V) Singleton premises and free distributive lattices

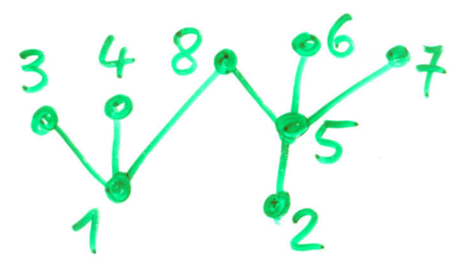
If Σ has all implications with singleton premises (thus $\{a\} \to B$) then the Σ-closed sets are essentially the order ideals of some poset (and conversely). Besides applications in e.g. scheduling, singleton premise implications occur in the calculation of finite modular lattices (free or not).

For instance, L below has poset $(P, \leq)$ of join irreducibles and base of lines Λ:

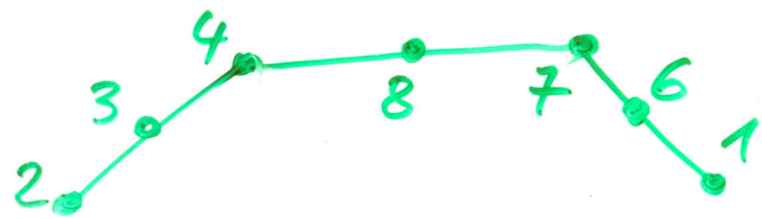
$L =$



$(P, \leq) =$



$\Lambda = \{\ell_1, \ell_2, \ell_3\}$



An order ideal X of $(P,\leq)$ is <u>Λ-closed</u> if

$$(\forall \ell \in \Lambda)\ |X \cap \ell| \geqslant 2 \Rightarrow \ell \subseteq X \quad (26)$$

It is clear that

$$\mathcal{L}(P,\leq,\Lambda) := \{X \subseteq P : X \text{ is } \Lambda\text{-closed ideal}\}$$

is a lattice, but less so that $a \mapsto J(a)$ yields an order isom. $\boxed{L \xrightarrow{\simeq} \mathcal{L}(P,\leq,\Lambda)}$ (Herrmann-Wild 1991). From this it is immediate that $\boxed{L \simeq \mathcal{C}(\Sigma)}$ where

$$\Sigma := \Sigma_{\leq} \cup \Sigma_{\Lambda} \text{ with}$$

$$\Sigma_{\leq} := \{\{a\} \to LC(a) : a \in P\}$$

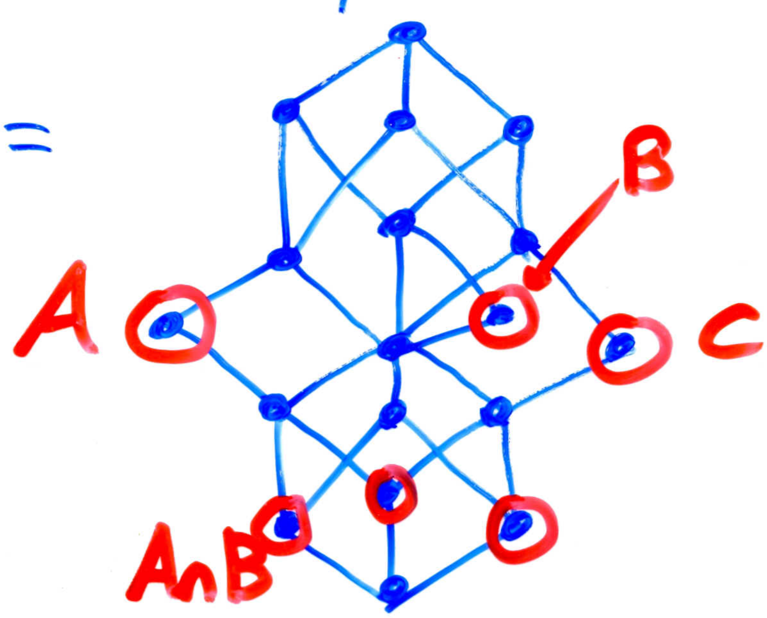
$$\Sigma_{\Lambda} := \{\{a,b\} \to \ell : \ell \in \Lambda,\ a \neq b \text{ in } \ell\}$$

This is particularly interesting for the calculation of (finite) <u>free</u> modular lattices.

V.1 New symmetries for FD(n)

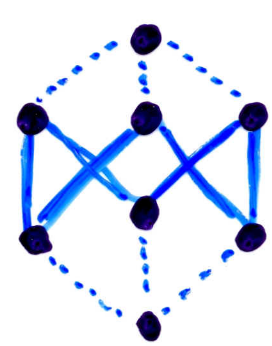
The cardinality of the free n-generated distributive lattice FD(n) equals the maximum number of sets obtainable from n sets by taking iterated unions and intersections at liberty. For instance:

FD(3) =



The smallest three sets that yield FD(3) are $A=\{1,2,3\}$, $B=\{1,3,4\}$, $C=\{2,3,5\}$.

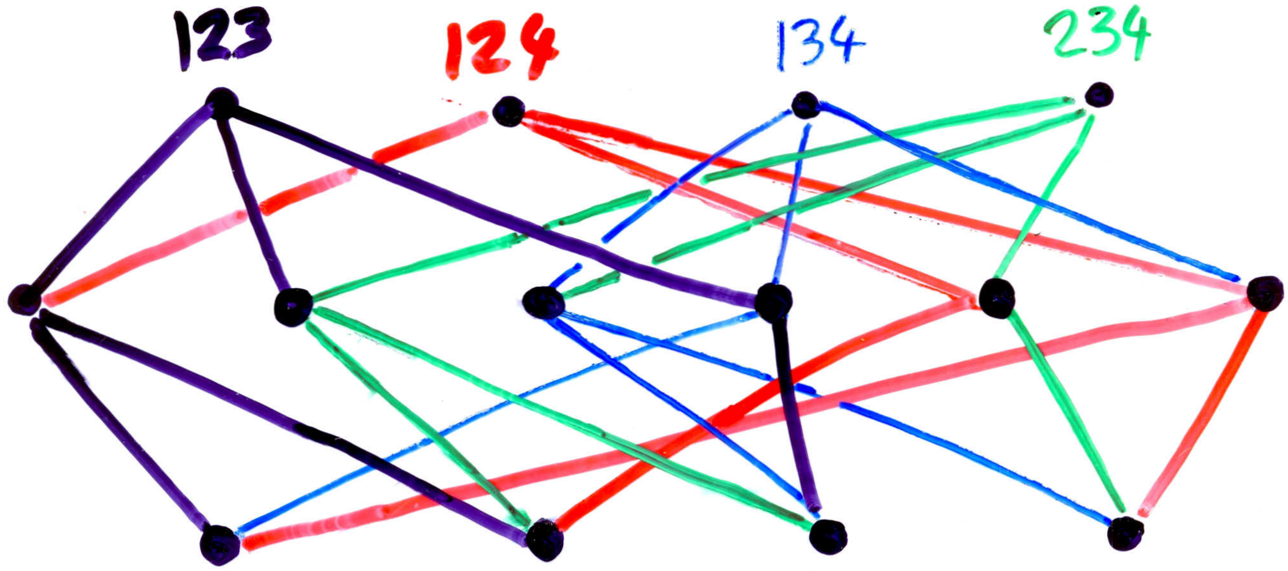
$J(FD(3)) =$



$=: P_3$

(the capped Boolean lattice B(3))

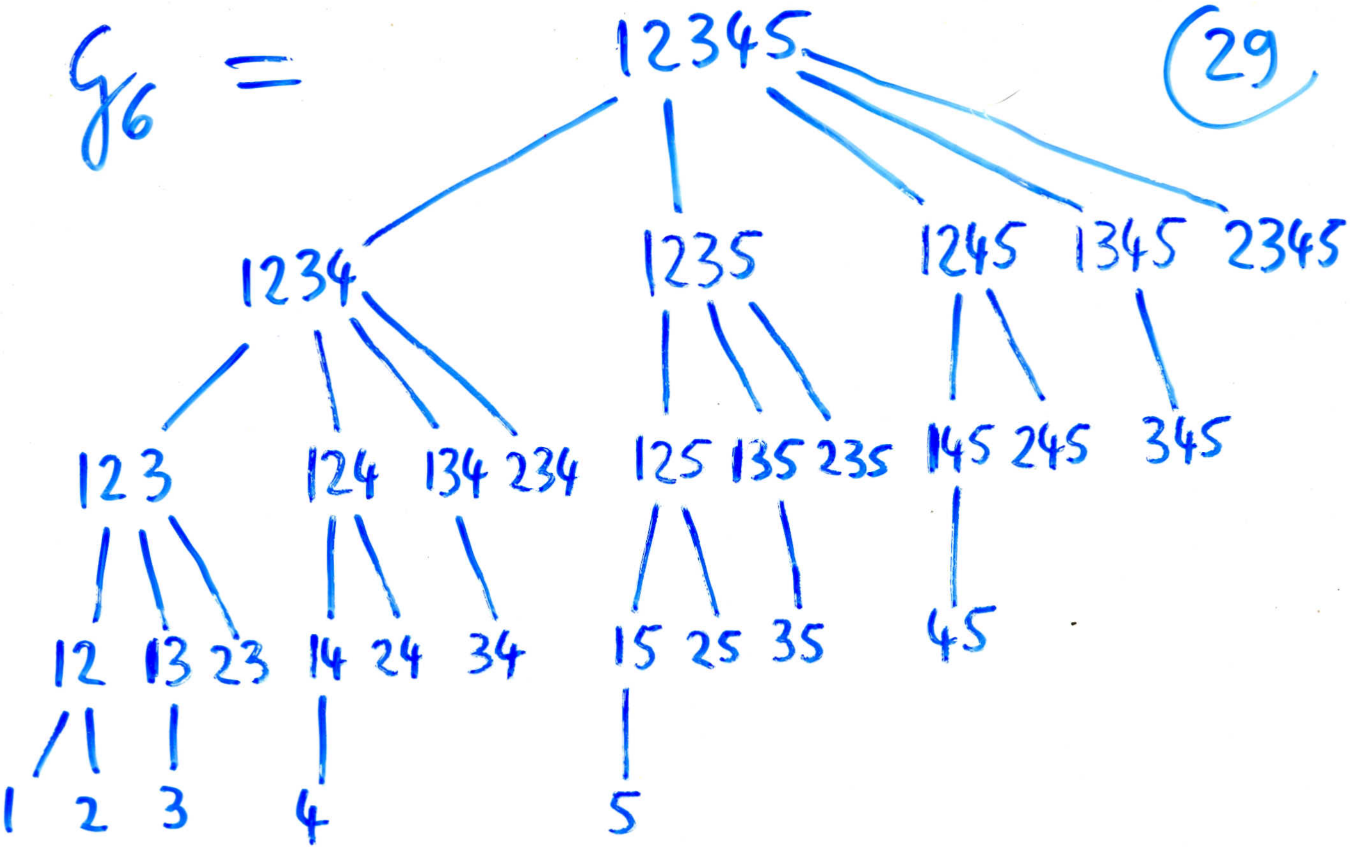
One can decompose $\mathcal{P}_n$ into an edge-disjoint union of n isomorphic trees (28)
$\mathcal{G}_n, \pi\mathcal{G}_n, \pi^2\mathcal{G}_n, \cdots, \pi^{n-1}\mathcal{G}_n$ where $\pi = (1, 2, \cdots, n)$
is a cyclic permutation. For instance



$$\mathcal{P}_4 = \mathcal{G}_4 \,\dot\cup\, \pi\mathcal{G}_4 \,\dot\cup\, \pi^2\mathcal{G}_4 \,\dot\cup\, \pi^3\mathcal{G}_4$$

As another example:

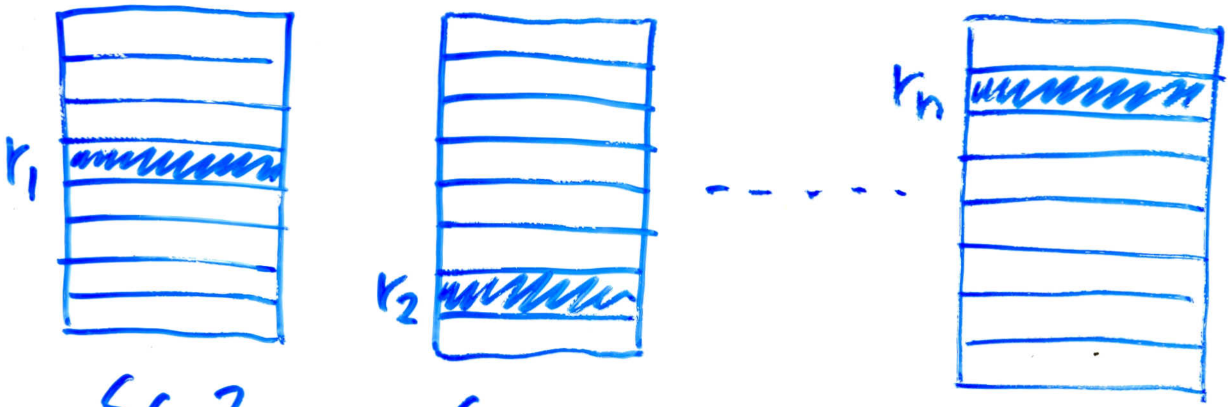
$\mathcal{G}_6 =$



One idea to calculate $|FD(n)| = |Id(\mathcal{P}_n)|$ by exploiting the symmetry of $\mathcal{P}_n$ is as follows. Define a <u>context</u> of $\mathcal{G}_n$ as

$\{\mathcal{G}_n\} :=$ Suitable collection of disjoint multivalued rows whose union is the set of order ideals of the tree $\mathcal{G}_n$

Then $|FD(n)|$ is obtained by "multiplying out" n isomorphic contexts: 30



$\{g_n\}$ $\{\pi g_n\}$ $\{\pi^{n-1} g_n\}$

In one version each context has 2^{n-1} multivalued rows, and thus each of the $(2^{n-1})^n$ transversals $\{r_1, r_2, \cdots, r_n\}$ contributes the summand $|r_1 \cap r_2 \cap \cdots \cap r_n|$ to $|FD(n)|$.

Fortunately $r_1 \cap r_2 \cap \cdots \cap r_n = \phi$ except for 2^n transversals $\{r_1, r_2, \cdots, r_n\}$. What's more, many of the latter are equivalent. It is hoped that $|FD(9)|$ yields to this approach. If not, turn to $|FM(n)|$ for suitable varieties M of modular lattices $(n \le 6)$.

VI) The transversal e-algorithm and frequent set mining

Let again Θ be a family of subsets of W. Observe that X is a Θ-noncover iff X^c is a transversal of Θ (since $X \not\supseteq Y \Leftrightarrow X^c \cap Y \neq \emptyset$).

The transversal e-algorithm is "the same" as the noncover n-algorithm except that it generates the transversals X directly (not as X^c). Observe:

$nn\cdots n$:= "at least one 0"
$ee\cdots e$:= "at least one 1"

The problem to find all inclusion-minimal transversals of a set system (hypergraph) $\mathcal{H}$ is well researched. The transversal e-algorithm looks interesting in what concerns minimum-cardinality transversals.

VI.1 Frequent set mining

Data mining e.g. comprises the theory of relational databases (closely linked to implications), and since 1993 the popular frequent set mining. The latter arose from an analysis of customer behaviour in a supermarket. The aim was to describe how often items were purchased together.

This led to the following abstract framework. Let W be a set of "items". A "transaction" is just a subset $R_i \subseteq W$. Fix a suitable threshold $t \in \mathbb{N}$. An itemset $X \subseteq W$ is called frequent if it is a subset of at least t transactions. Formally, if

$$\operatorname{supp}(X) := \{ i \in I : X \subseteq R_i \}$$

then "frequent" means that $|\operatorname{supp}(X)| \geq t$.

Obviously the family $\mathcal{SC}$ of frequent sets constitutes a simplicial complex, i.e. is closed under taking subsets. Early attempts (A priori algorithm and its variants) generated $\mathcal{SC}$ one by one "level-wise". For large $\mathcal{SC}$ this is not feasible.

Fortunately it often suffices to know some prominent features of $\mathcal{SC}$; we sketch three of them:

(A) The maximal members (facets) of $\mathcal{SC}$

(B) The face-numbers

$$f_k := |\{X \in \mathcal{SC} : |X| = k\}|$$

(C) The closed members $Y \in \mathcal{SC}$ in the sense that

$$Y \subsetneqq Y' \Rightarrow supp(Y') \subsetneqq supp(Y)$$

It is easily seen that the closed $Y \in \mathcal{SC}$ are exactly the frequent sets in the clos. system gen. by R_i $(i \in I)$.

As to A, getting the facets of a simpl. complex has generally many applications. Once they are given, BDD's compute $|SC|$ fast.

As to B, as previously argued, BDD's cannot count fixed-cardinality models. Here's how it works. Let

$$\mathcal{H} := \{F_1^c, F_2^c, \ldots, F_h^c\}$$

be the family of the complements of the provided facets of SC. For any $X \subseteq W$:

$$X \notin SC \Leftrightarrow X \text{ is a transversal of } \mathcal{H}$$

For each fixed $k \in \{1, 2, \ldots, w\}$ the number τ_k of k-element $\mathcal{H}$-transversals can be determined fast with the e-algorithm. One hence gets the face-numbers f_k as

$$\boxed{f_k = \binom{w}{k} - \tau_k}$$

As to (C), how does one generally calculate, from given sets R_i $(i \in I)$, the closure system (35)

$$\mathcal{C} := \left\{ \bigcap_{j \in J} R_j : J \subseteq I \right\} \;?$$

(In our frequent set mining application only those $Y \in \mathcal{C}$ with $supp(Y) \geq t$ are required.)

For small $\mathcal{C}$ one can use Ganter's algorithm. Otherwise strive for an implicational base Σ of $\mathcal{C}$ and apply the implication n-algorithm to it.