Computing the D-base of Finite Closure Systems

Algo Seminar, GREYC, Caen

Feb. 2024

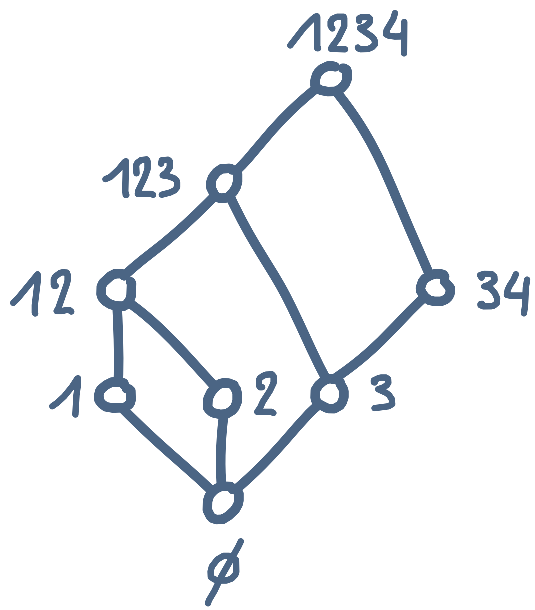
Simon Vilmin — LIS, Aix-Marseille Université

Joint work with:

Kira Adaricheva — Dep. of Mathematics, Hofstra University, USA

Lhouari Nourine — LIMOS, Université Clermont Auvergne

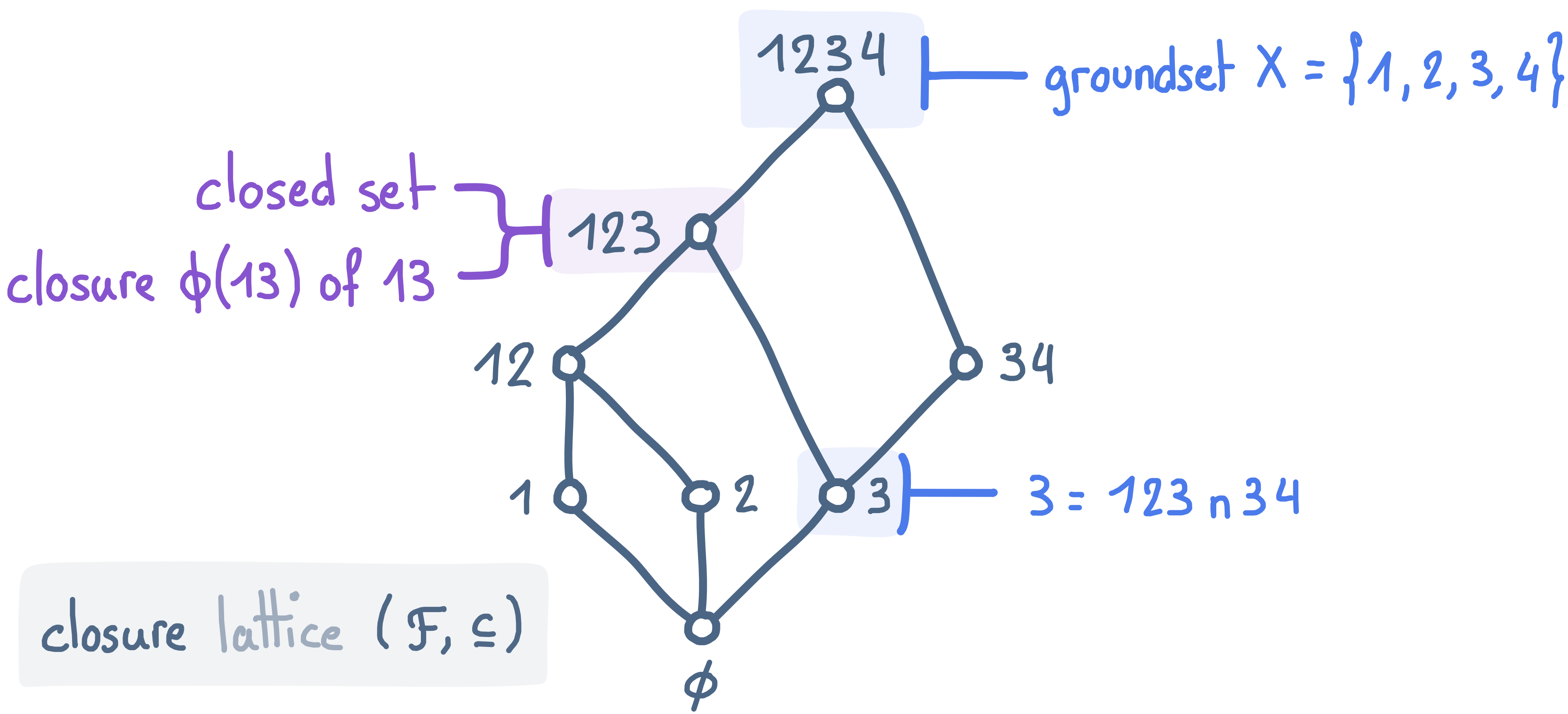
Closure systems : what, how, why



Finite closure systems?

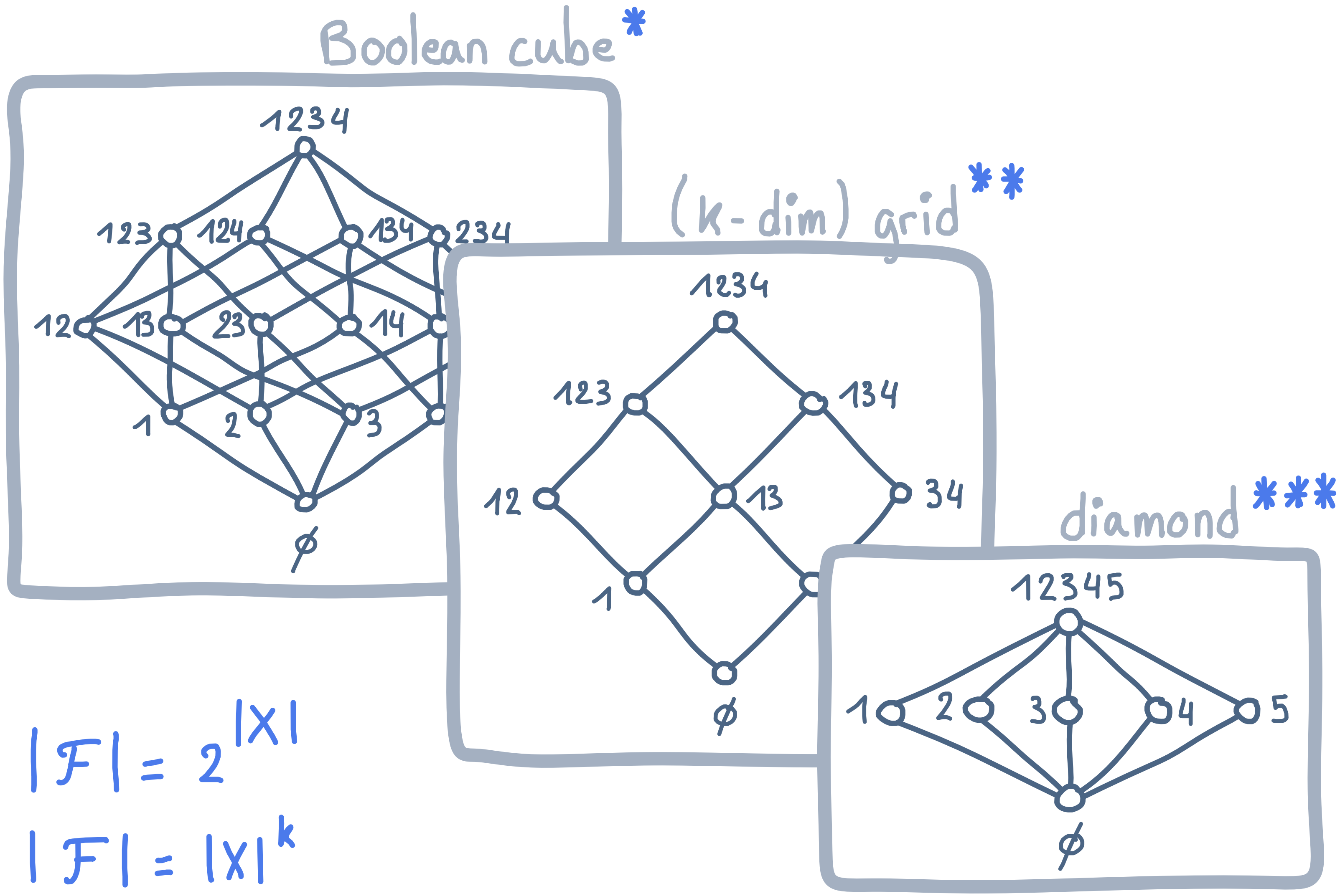
X finite groundset, $\mathcal{F} \subseteq 2^X$

DEF (closure system): set system $\boxed{(X, \mathcal{F})}$ s.t. $X \in \mathcal{F}$ and $F_1 \cap F_2 \in \mathcal{F}$ for all $F_1, F_2 \in \mathcal{F}$



closure lattice $(\mathcal{F}, \subseteq)$

Closure systems : some examples

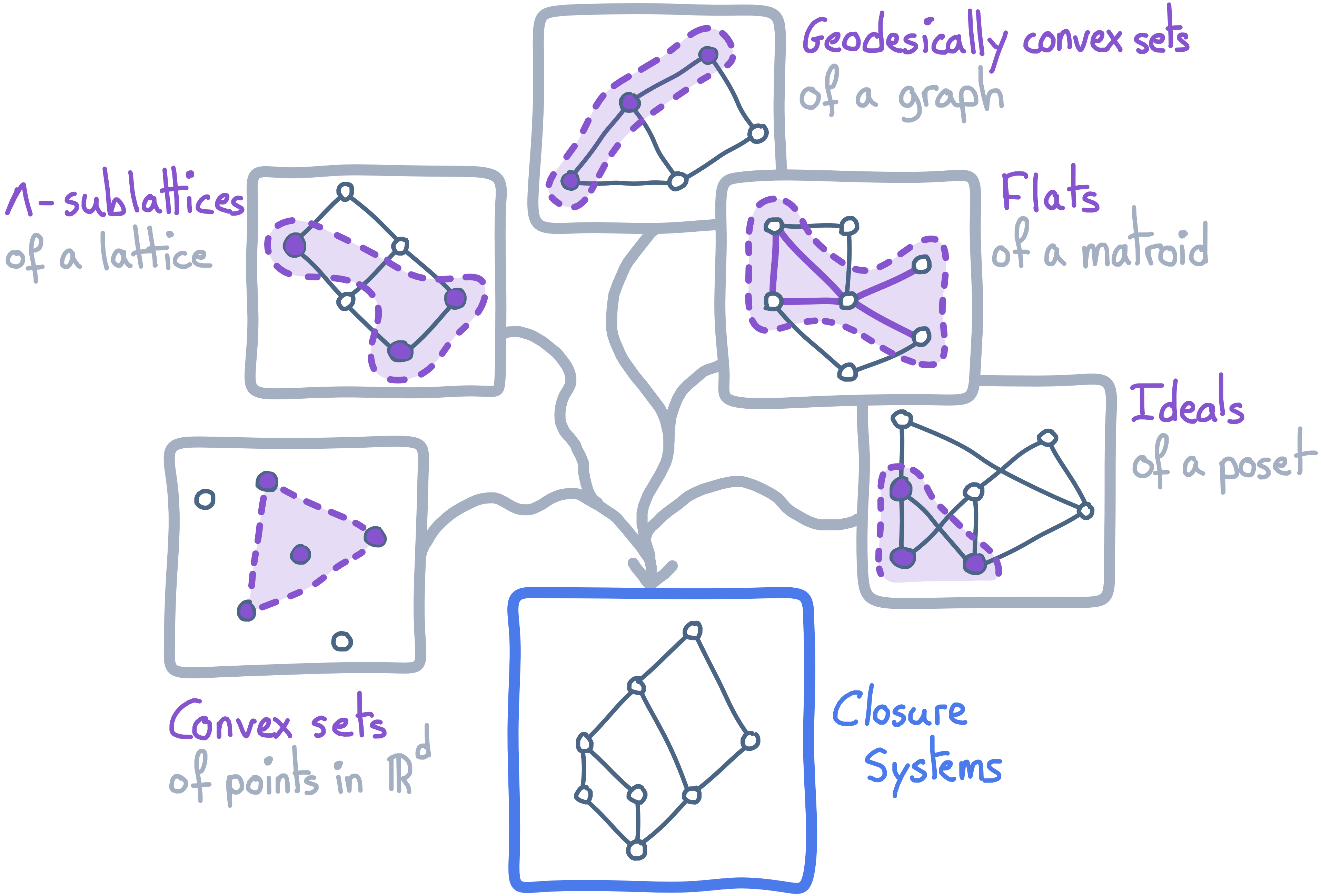


$*\ |\mathcal{F}| = 2^{|X|}$

$**\ |\mathcal{F}| = |X|^k$

$***\ |\mathcal{F}| = |X| + 2$

(some) sources of closure systems



Representations of closure systems

Closure systems

- arise from numerous objects and fields
- often not given explicitly
- have HUGE size and may be hard to understand

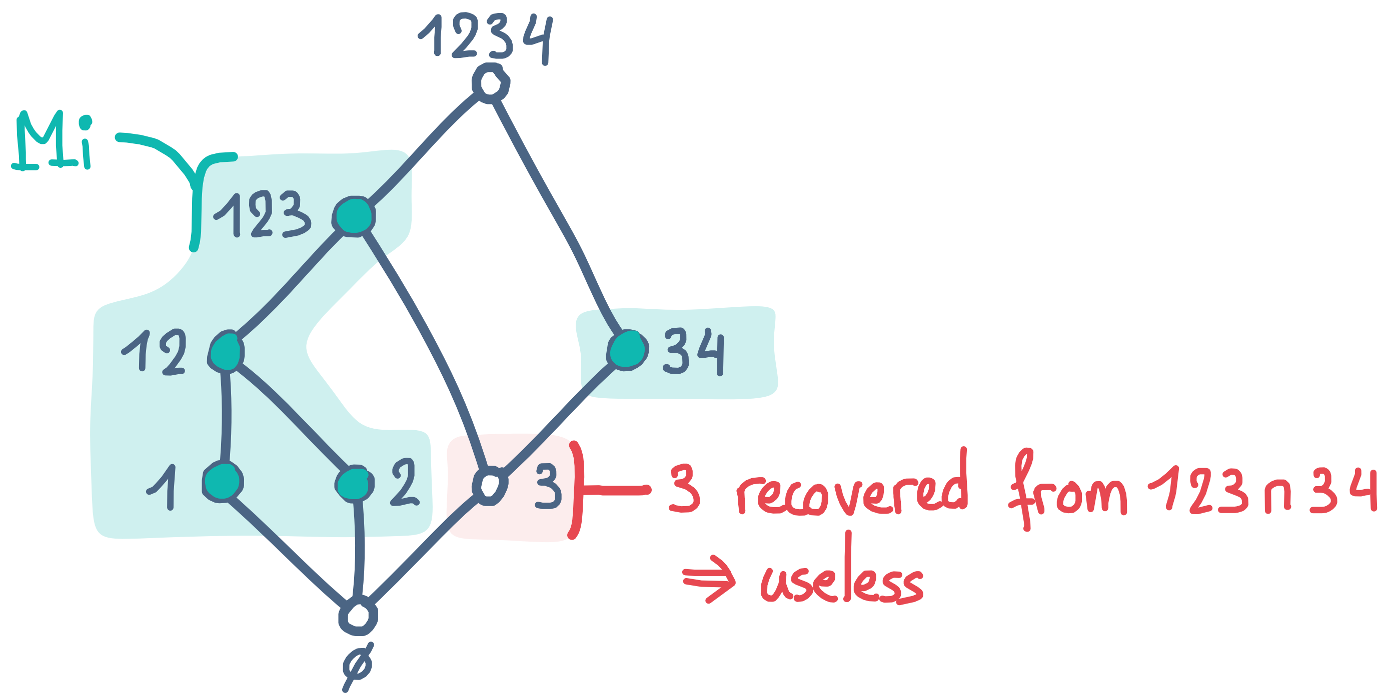
IDEA : work with implicit representations

Among representations, two VIPs :

- meet-irreducible closed sets
- implicational bases

] encode any closure system!

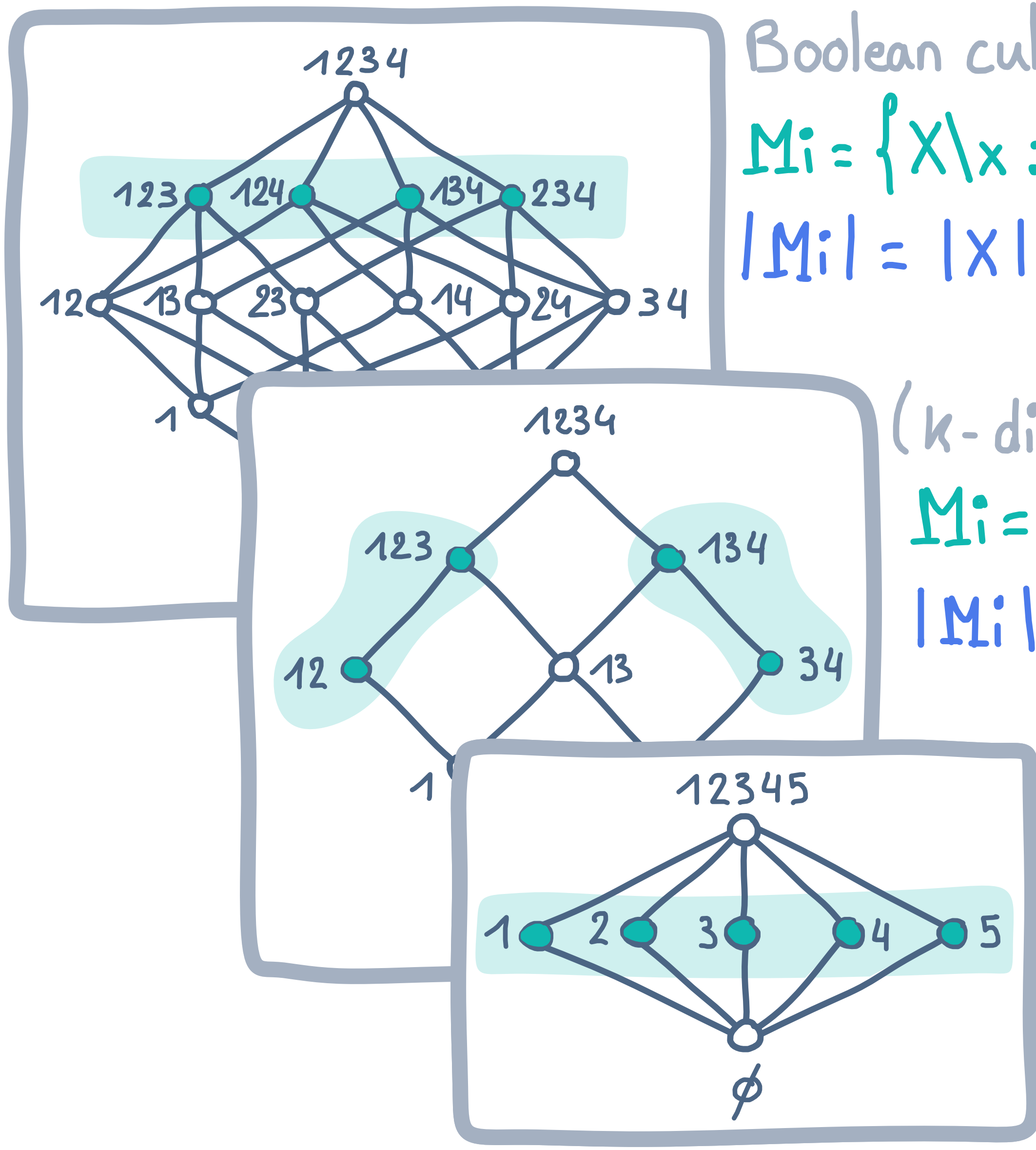
IDEA : discard useless closed sets, keep the essential ones



DEF (meet-irreducible) : closed set $M \neq X$ not obtained by intersecting other closed sets. We put

$$\mathcal{M}i = \{ M \in \mathcal{F} : M \text{ meet-irreducible} \}$$

Meet-irreducible closed sets: some examples



Boolean cube

$Mi = \{X \setminus x : x \in X\}$

$|Mi| = |X| \ll 2^{|X|} = |\mathcal{F}|$

(k-dim) grid

$Mi = \{X \setminus \phi(x) : x \in X\}$

$|Mi| = |X| < |X|^k = |\mathcal{F}|$

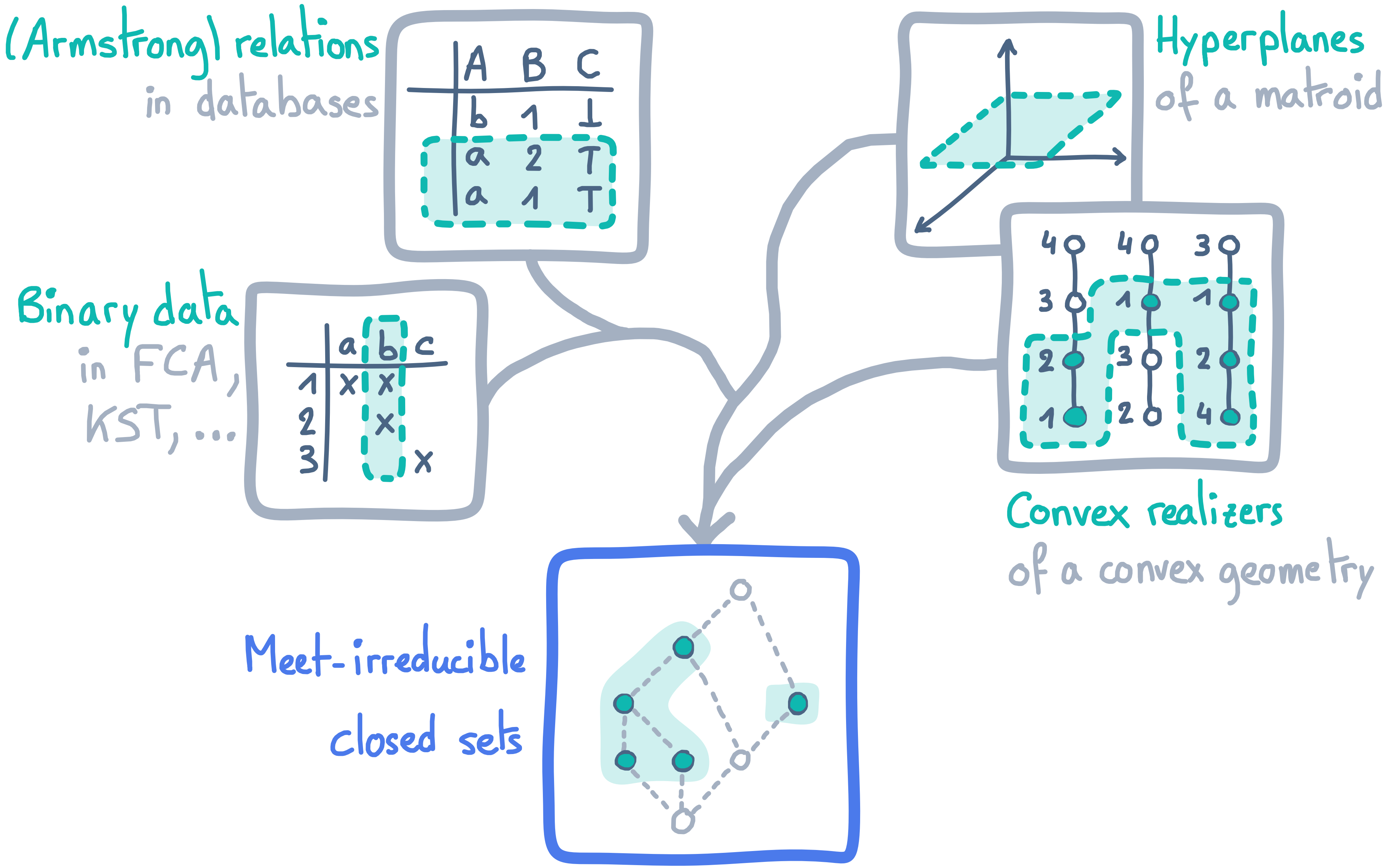
diamond

$Mi = \{\{x\} : x \in X\}$

$|Mi| = |X| \simeq |X| + 2 = |\mathcal{F}|$

⚠ : in general, $|Mi| = O(2^{|X|})$

(further) sources of meet-irreducible closed sets



Implications, implicational bases (IBs)

(or $A \to B$, $B \subseteq X$)

DEF (syntax):

- implication: statement $\boxed{A \to b}$ with $A \subseteq X$, $b \in X$
- implicational base (IB): pair (X, Σ) where Σ set of implications over X

DEF (semantic): a set $F \subseteq X$ is closed w.r.t. (X, Σ) if for all $A \to b \in \Sigma$, $\boxed{A \subseteq F \text{ implies } b \in F}$

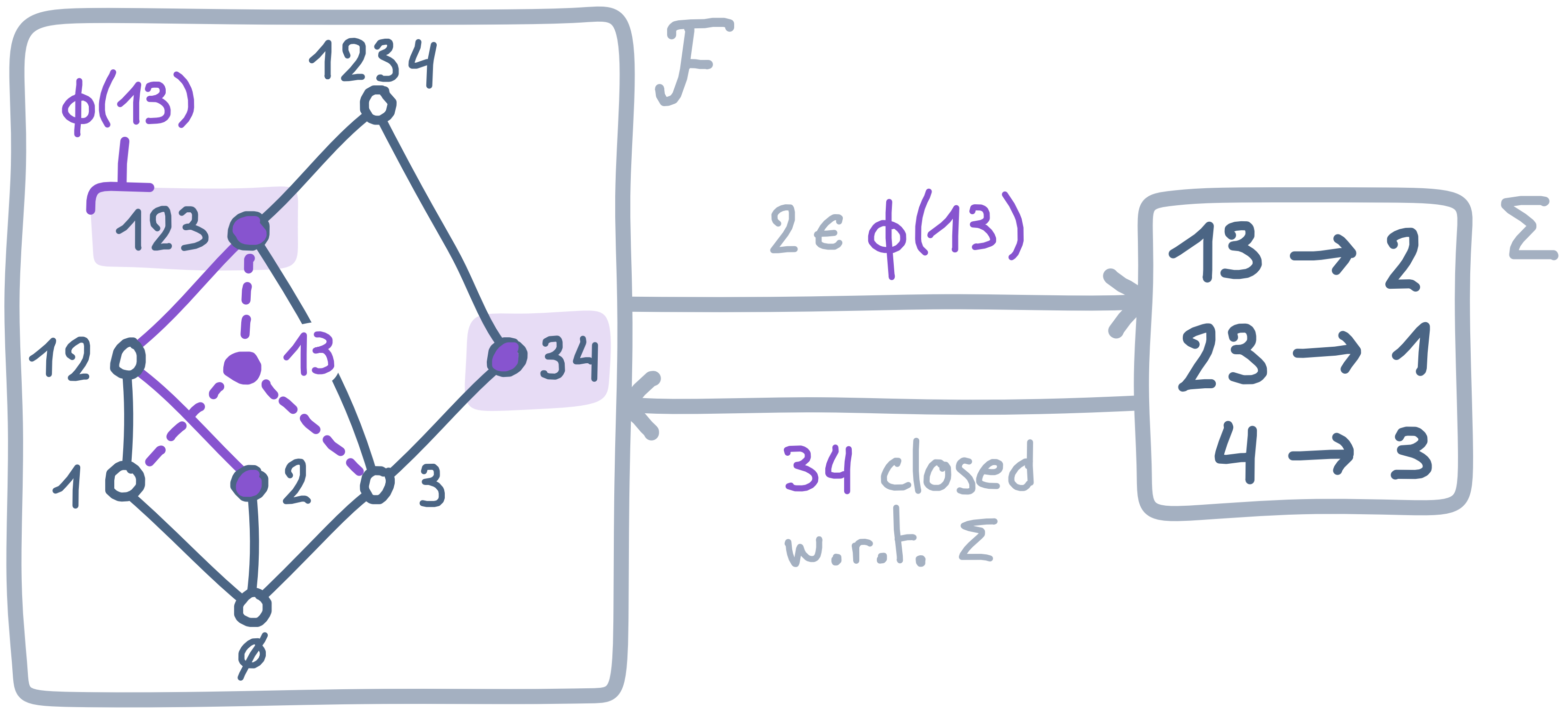
"If we have A, we have b"

$$\Sigma = \begin{bmatrix} 4 \to 3 \\ 13 \to 2 \\ 23 \to 1 \end{bmatrix}$$

14 ✗

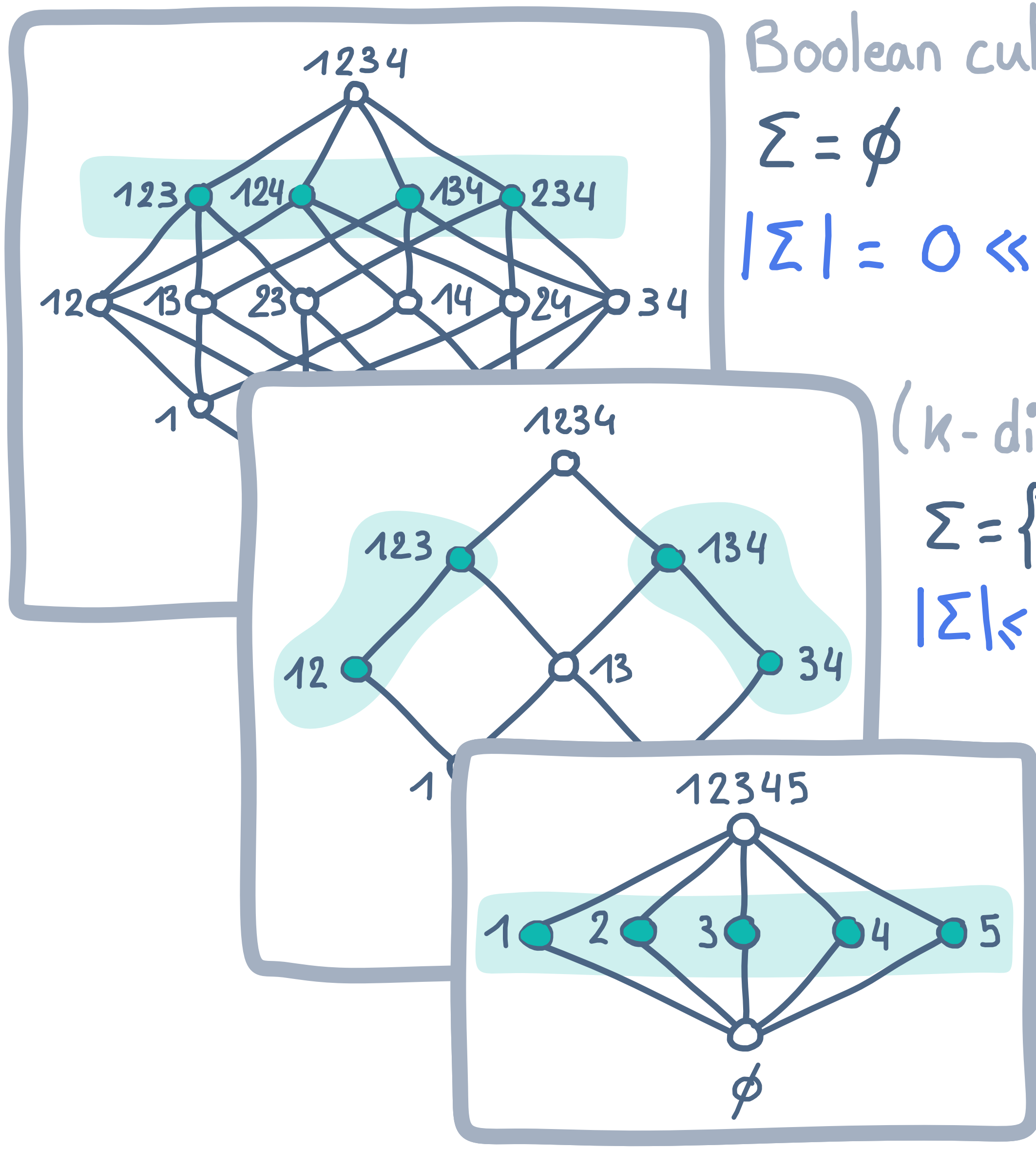
12 ✓

Connections with closure systems



THM (folklore): there is a correspondence between closure systems and implicational bases

Implicational bases: some examples



Boolean cube

$\Sigma = \emptyset$

$|\Sigma| = 0 \ll 2^{|X|} = |\mathcal{F}|$

(k-dim) grid

$\Sigma = \{x \to y : x, y \in X, y \in \phi(x)\}$

$|\Sigma| \leq |X|^2 \ll |X|^k = |\mathcal{F}|$

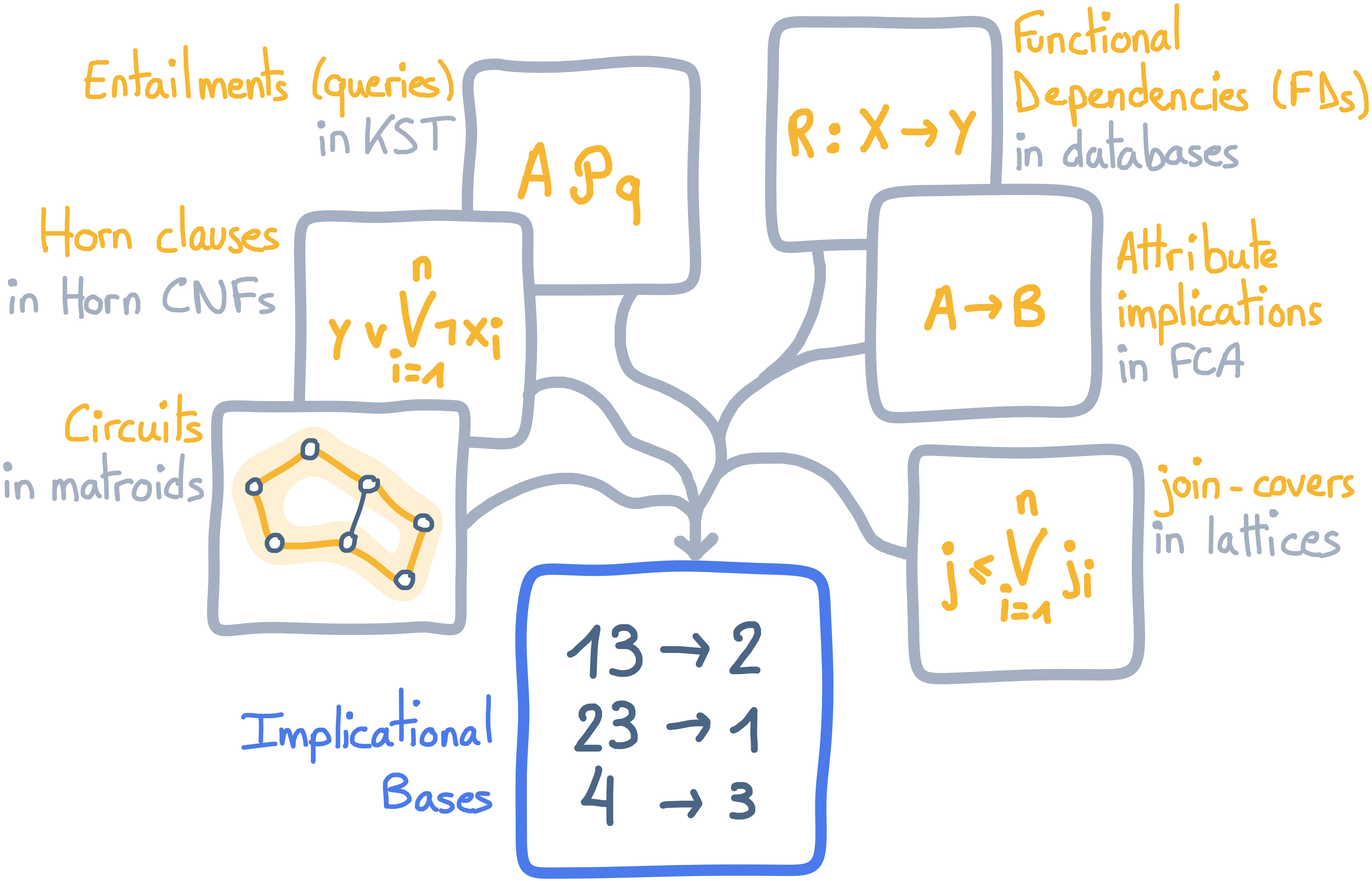
diamond

$\Sigma = \{xy \to z : x, y, z \in X\}$

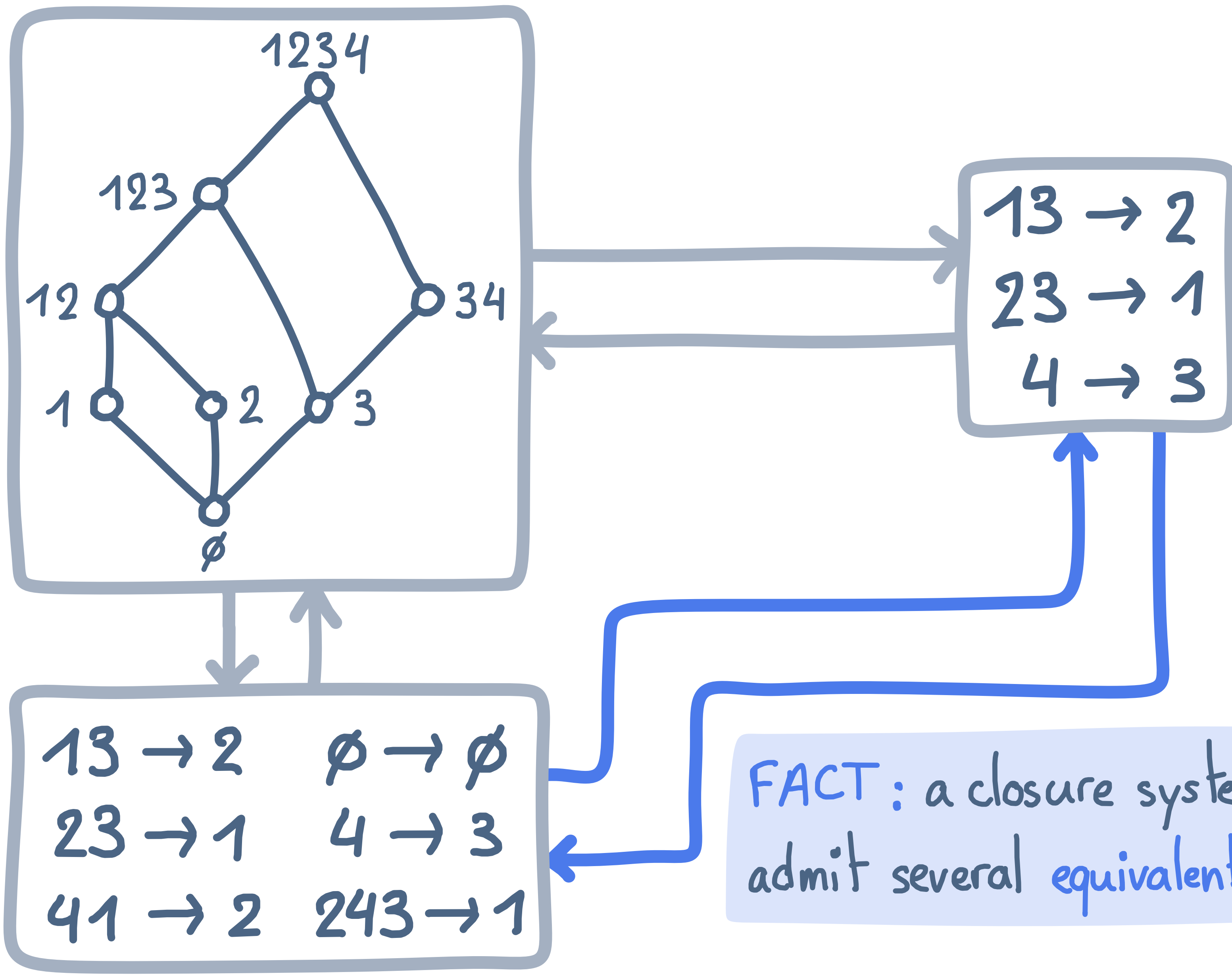
$|\Sigma| = |X|^3 \geq |X| + 2 = |\mathcal{F}|$

⚠ : in general, $|\Sigma| = O(2^{|X|})$

(again) sources of IBs



One closure system, many IBs



FACT: a closure system may admit several equivalent IBs

A glimpse into the galaxy of IBs

$\boxed{\Sigma_c}$

Canonical / Duquenne-Guigues

- uses pseudo-closed sets
- unique
- minimum

$\boxed{\Sigma_{cd}}$

Canonical direct

- uses minimal generators
- unique
- direct ($\approx>$ fast closure)

$\boxed{\Sigma_D}$

D-base

- uses D-generators
- unique
- ordered direct

$\boxed{\Sigma_{min}}$

Minimum

minimum number of implications

$\boxed{\Sigma_{op}}$

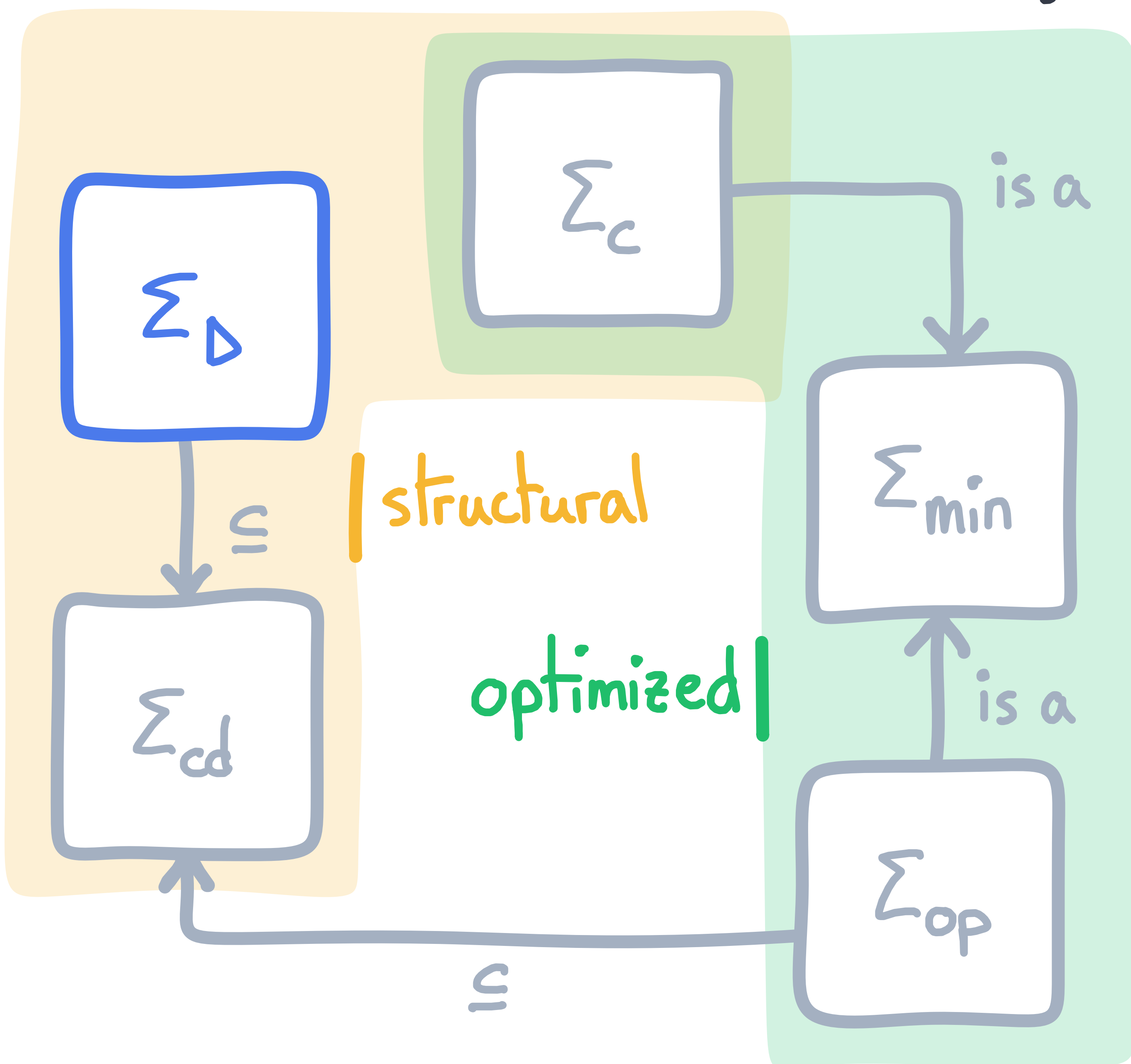
Optimum

minimum total size

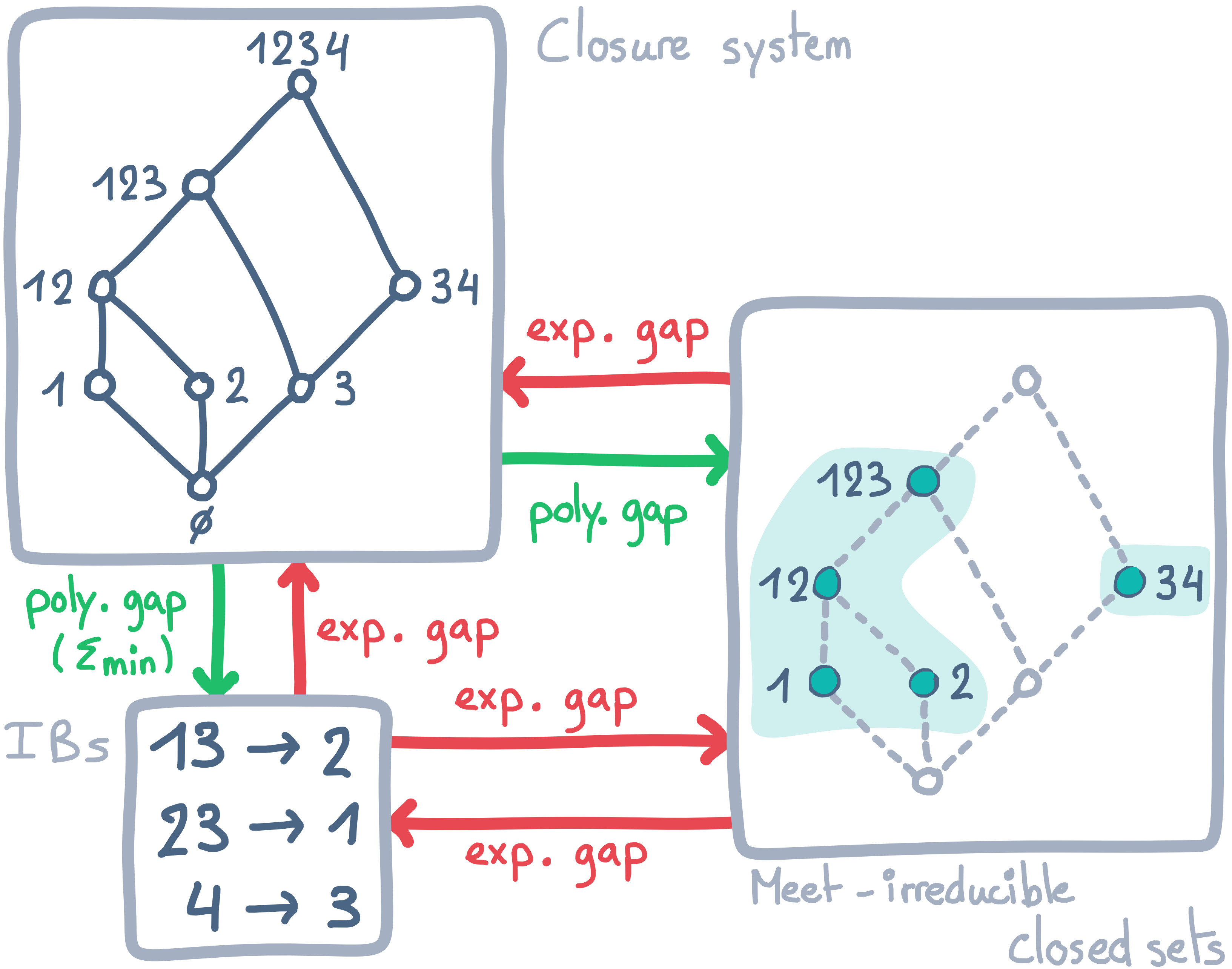
$(\Sigma |A|+1, \ A \to b \in \Sigma)$

see Wild, 17

A glimpse into the galaxy of IBs



Same information, different POVs



changing representation : a natural problem

IDEA : given a representation, find another one

EX : given an $\mathbb{B}$, find a minimum/optimum $\mathbb{B}$

- Horn minimization, optimum covers of FDs, ...

EX : given $\mathcal{M}_i$, find an $\mathbb{B}$ (minimum, canonical direct, ...)

- find FDs / implications / rules in (binary) data, ...

EX : given an $\mathbb{B}$, find $\mathcal{M}_i$

- build knowledge space from queries, find characteristic models of Horn CNF, ...

- given an $\mathbb{B}$, find a "smaller" $\mathbb{B}$
  "optimize an $\mathbb{B}$"
  output size bounded by input size

- given an $\mathbb{B}$, find $\mathcal{M}_i$
  "enumerate the sets in $\mathcal{M}_i$"
  output size $\boxed{\text{exponential}}$ w.r.t. input size (in general)
  even "good" algorithms will have exponential running time ...

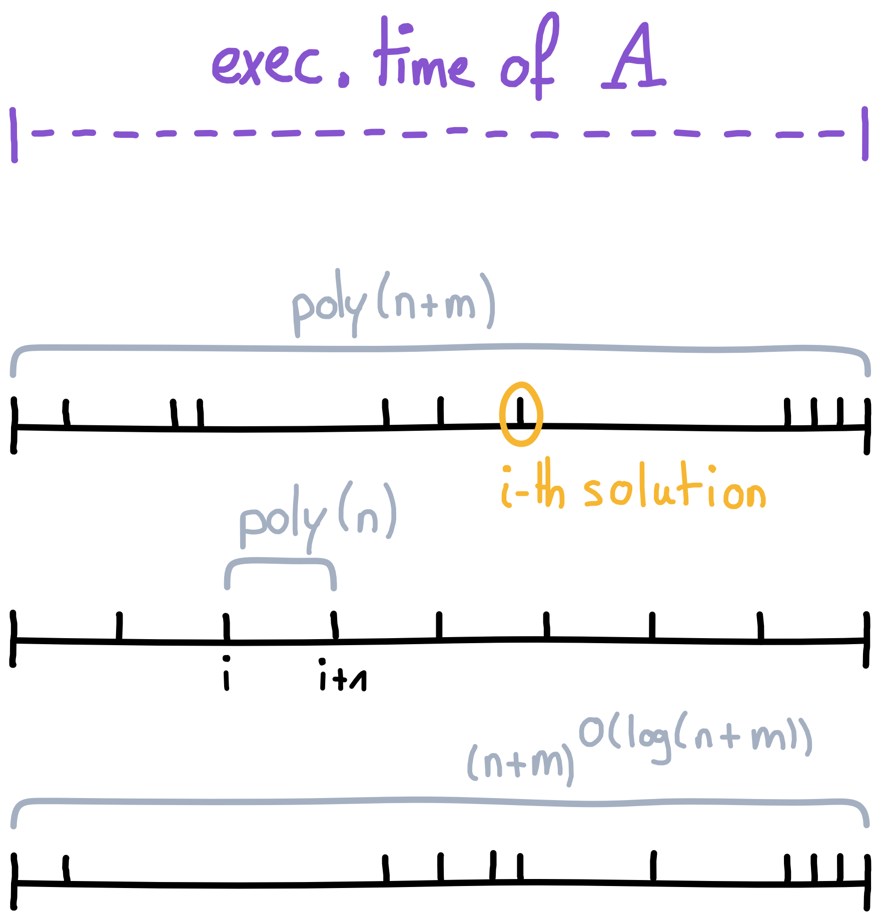
IDEA: output-sensitive complexity Johnson et al., 88

Enumeration: output-sensitive complexity

Each of size poly(x)

Enumeration task: with input x, list a set of solutions $R(x)$

Enumeration algorithm A
input x of size n
output $R(x)$ of size m



Output polynomial time

polynomial delay

Output quasi-polynomial time

Some important results

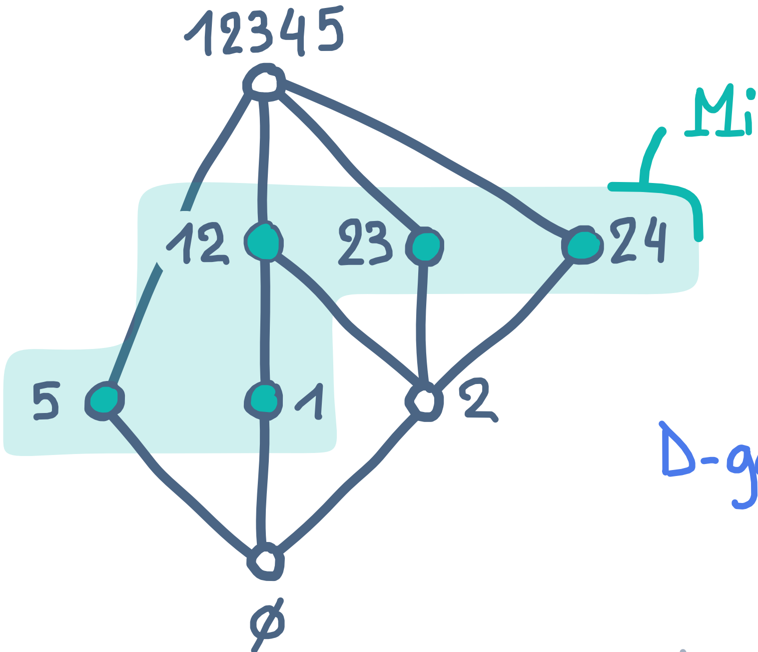
given (with X)	find	Complexity	
any Σ	Σ_c, Σ_{min}	poly	Maier, 80 Duquenne, Guigues, 86
Σ_c, Σ_{min}, Σ_{cd}, Σ_D	Σ_{op}	NP-C	Ausiello et al., 86
$\mathcal{M}_i$	Σ_{cd}	quasi-poly *	Khardon, 95
$\mathcal{M}_i$	Σ_{min}	Open	Wild, 17 V., Nourine, 23

* equiv. to hypergraph dualization !

The D-base: enumeration algorithms

Among minimal generators,
keep the "binary-closure" minimal

Brand new toy example



Binary implications Σ_b

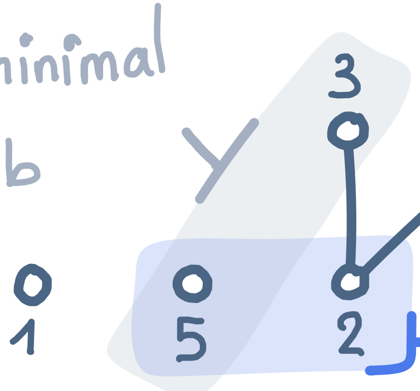
$$\Sigma = \left[\begin{array}{ll} 4 \to 2, & 25 \to 3, \\ 3 \to 2, & 34 \to 1, \\ 25 \to 4, & 14 \to 3, \\ 15 \to 4, & 13 \to 5, \\ 35 \to 4 & \end{array}\right]$$

D-generator of 4

minimal generators of 4

(no subset of 35 gives 4)

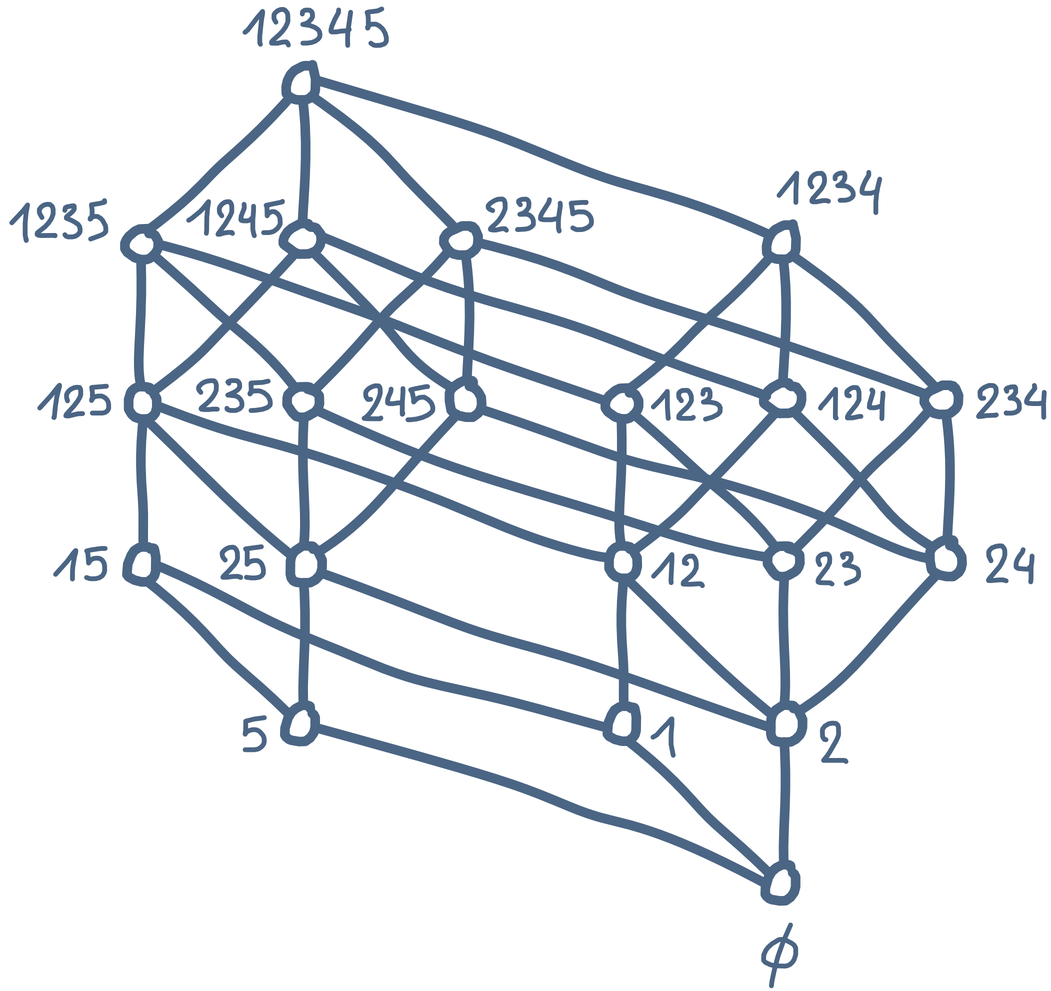
35 not minimal w.r.t. Σ_b



Σ_b as a poset

25 minimal w.r.t. $\Sigma_b \Rightarrow$ D-generator of 4

Binary implications



DEF: Binary implication: implication $a \to b$, $a, b \in X$

Closure system $(X, \mathcal{F}_b)$
and operator ϕ_b
associated with (X, Σ_b)
$\Sigma_b = \{3 \to 2, 4 \to 2\}$

Birkhoff, 37

THM (Birkhoff): a closure system (lattice) is distributive iff it is isomorphic to the closure system of some binary IB

Minimal Generators

DEF (minimal generator): $A \subseteq X$ minimal generator of x if $\subseteq$-minimal subset satisfying $\boxed{A \to x}$

- "minimal ways of deriving x"
- Circuits of a matroid
- prime implicate
- LHS-Minimal FD

$$\Sigma = \begin{bmatrix} 4 \to 2, & 25 \to 3, \\ 3 \to 2, & 34 \to 1, \\ 25 \to 4, & 14 \to 3, \\ 15 \to 4, & 13 \to 5, \\ 35 \to 4 & \end{bmatrix}$$

$35 \to 2$ ✗ $5 \to 2$

$234 \to 1$ ✗ $34 \to 1$

$15 \to 3$ ✓

D-generators, D-base Adaricheva et al., 13

DEF (D-generator, D-base):

- D-generators of x: among minimal generators of x, those with $\subseteq$-minimal closure w.r.t. binary implications
- THE D-base (X, Σ_D) of a closure system:

$$\Sigma_b + \{A \to x : x \in X, A \text{ D-gen of } x\}$$

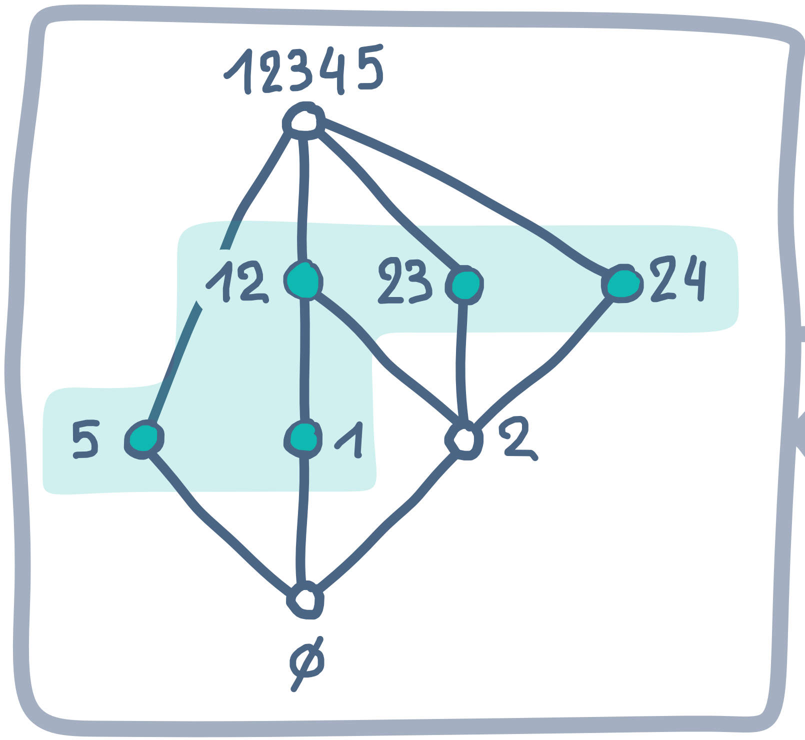
$$\Sigma = \begin{bmatrix} 4 \to 2, & 25 \to 3, \\ 3 \to 2, & 34 \to 1, \\ 25 \to 4, & 14 \to 3, \\ 15 \to 4, & 13 \to 5, \\ 35 \to 4 & \end{bmatrix}$$

$35 \to 1$ ✗

$45 \to 1$ ✗

$25 \to 1$

On our example



Closure system

D-base (X,Σ_Δ)

$$\overbrace{\{4 \to 2, 3 \to 2\}}^{\Sigma_b} + \begin{bmatrix} 34 \to 1, & 25 \to 1, & 15 \to 2, \\ 13 \to 2, & 15 \to 3, & 14 \to 3, \\ 25 \to 3, & 15 \to 4, & 25 \to 4, \\ 13 \to 5, & 34 \to 5, & 14 \to 5 \end{bmatrix}$$

Motivation

Theoretical / algorithmic properties :

- convey structural information of closure systems
- ordered direct (fast forward chaining)
- much smaller than the set of all minimal generators

Practical uses :

- seabreeze forecast Adaricheva et al., 23
- stomach cancer risk estimation Nation et al., 21

How **hard** is it to change the representation ?

↑ more generally

Recover the D-base to enjoy its properties

↓ more precisely

D-base from Mi (DB-M): given Mi, find (X, Σ_D)

D-base from Σ (DB-IB): given (X, Σ), find (X, Σ_D)

D-base from $\mathcal{M}i$ (DB-M): given $\mathcal{M}i$, find (X, Σ_D)

ANV, 23+

DB-M can be solved in output quasi-polynomial time

Our approach: dualization

Existing work:

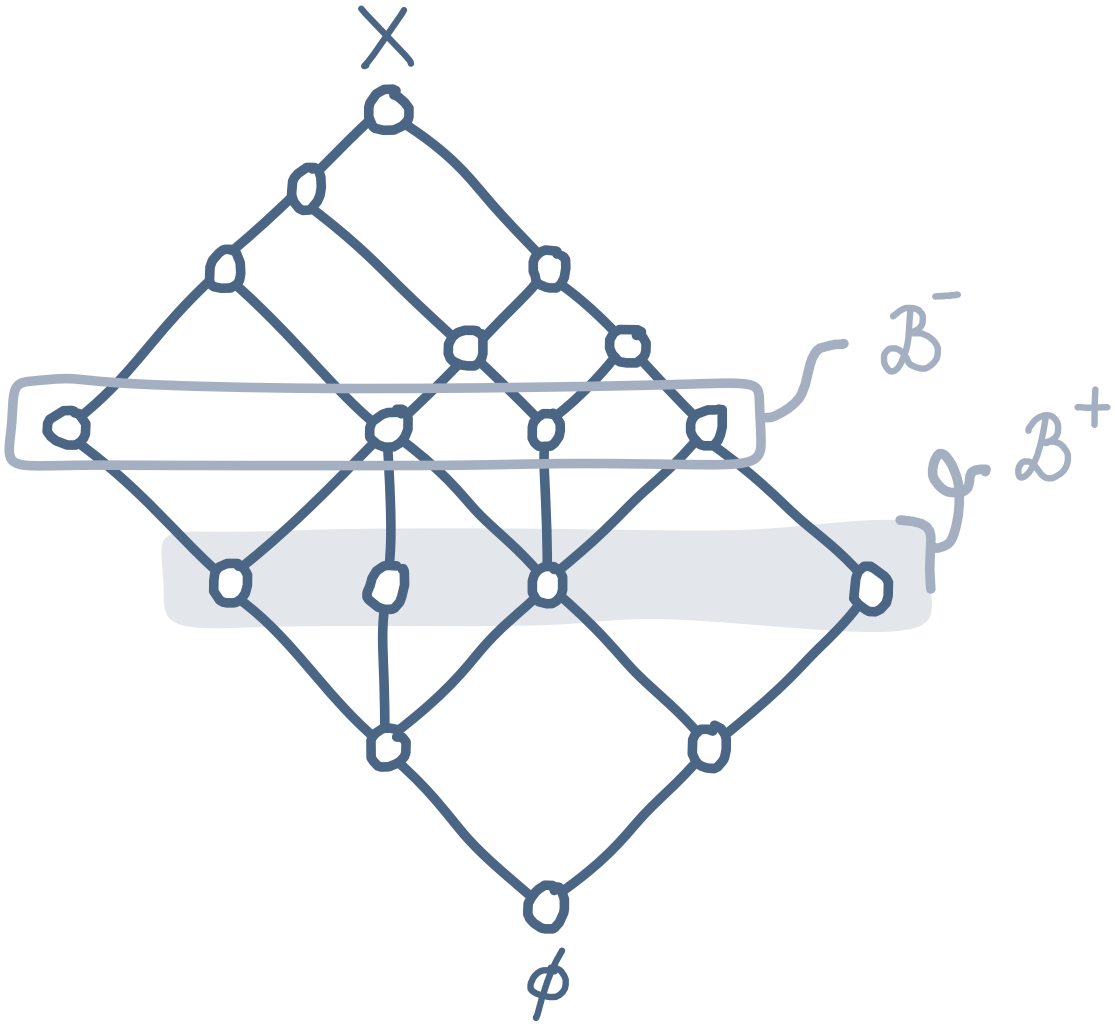
- algorithm based on Hypergraph dualization Adaricheva, Nation, 17
  produces (possibly large) superset of D-base

IDEA: D-base relies on Σ_b

Σ_b defines a distributive closure system

$\Rightarrow$ use dualization in distributive closure systems

Dualization (with Σ)

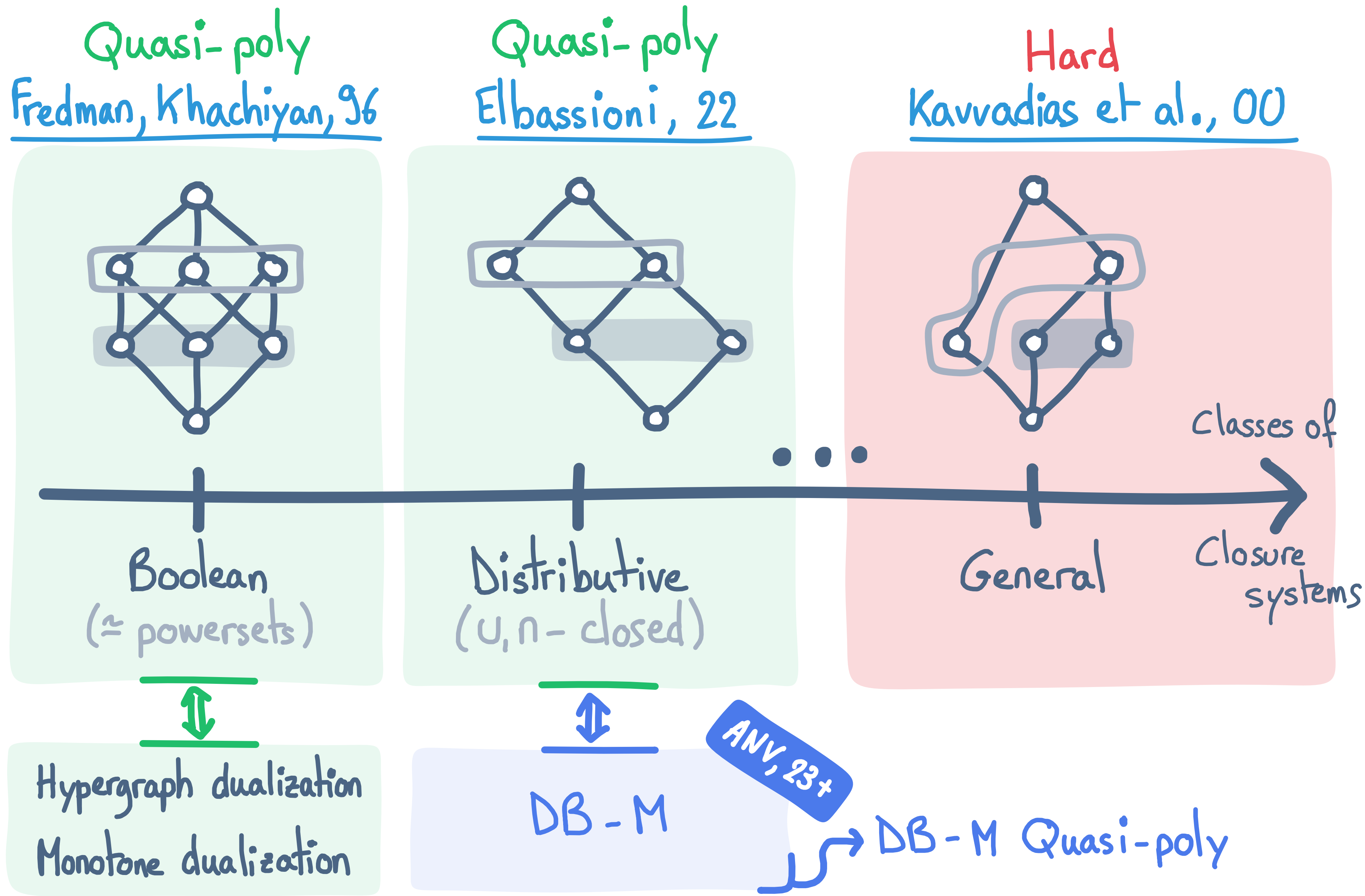


$\mathcal{B}^-$ and $\mathcal{B}^+$ are dual:

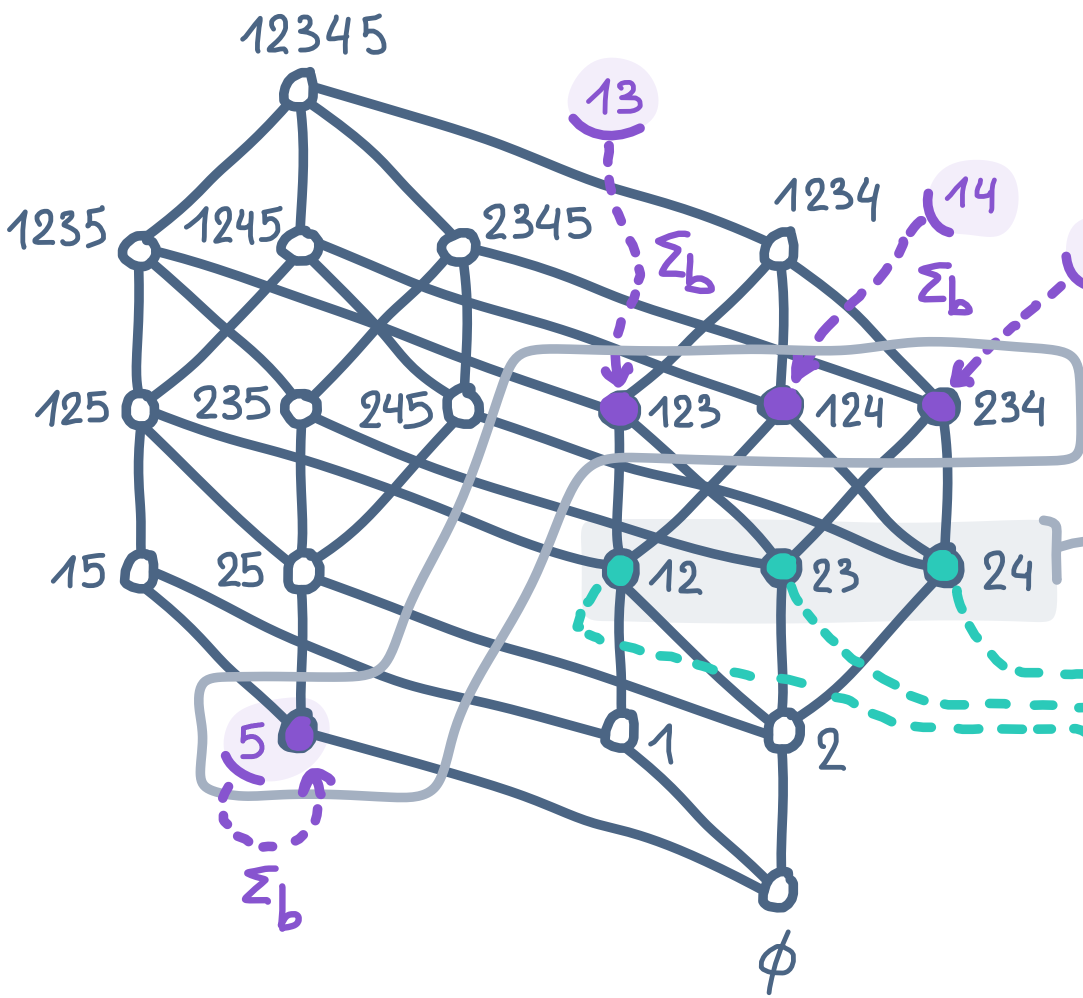
- $\downarrow\mathcal{B}^+ \cup \uparrow\mathcal{B}^- = \mathcal{F}$
- $\downarrow\mathcal{B}^+ \cap \uparrow\mathcal{B}^- = \emptyset$

Dualization: with (X,Σ) and antichain $\mathcal{B}^+$, find antichain $\mathcal{B}^-$

Dualization complexity (with Σ) and DB-M

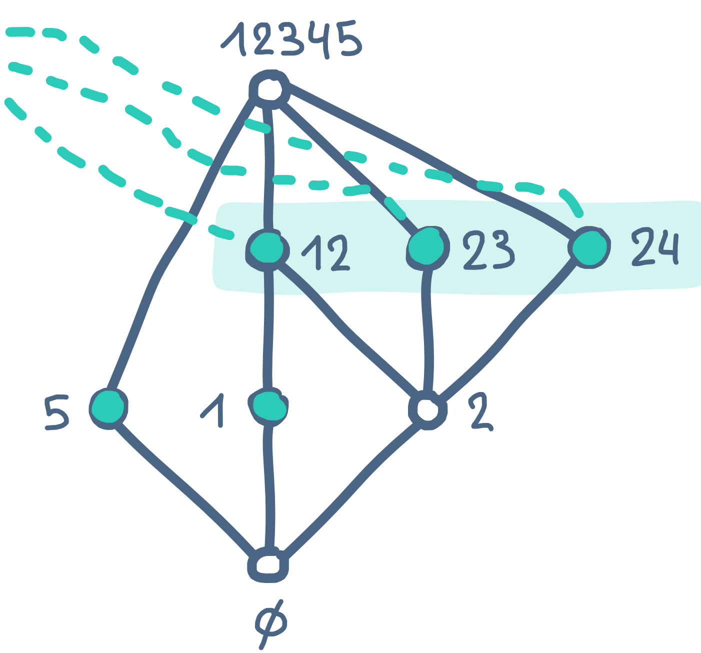


Intuition: DB-M ≈ Dualization Distr.



② $\mathcal{B}^-$ obtained from 5 and D-generators of 5

③ $\mathcal{B}^+$ = $\subseteq$-max irreducibles without 5



① distributive closure system of

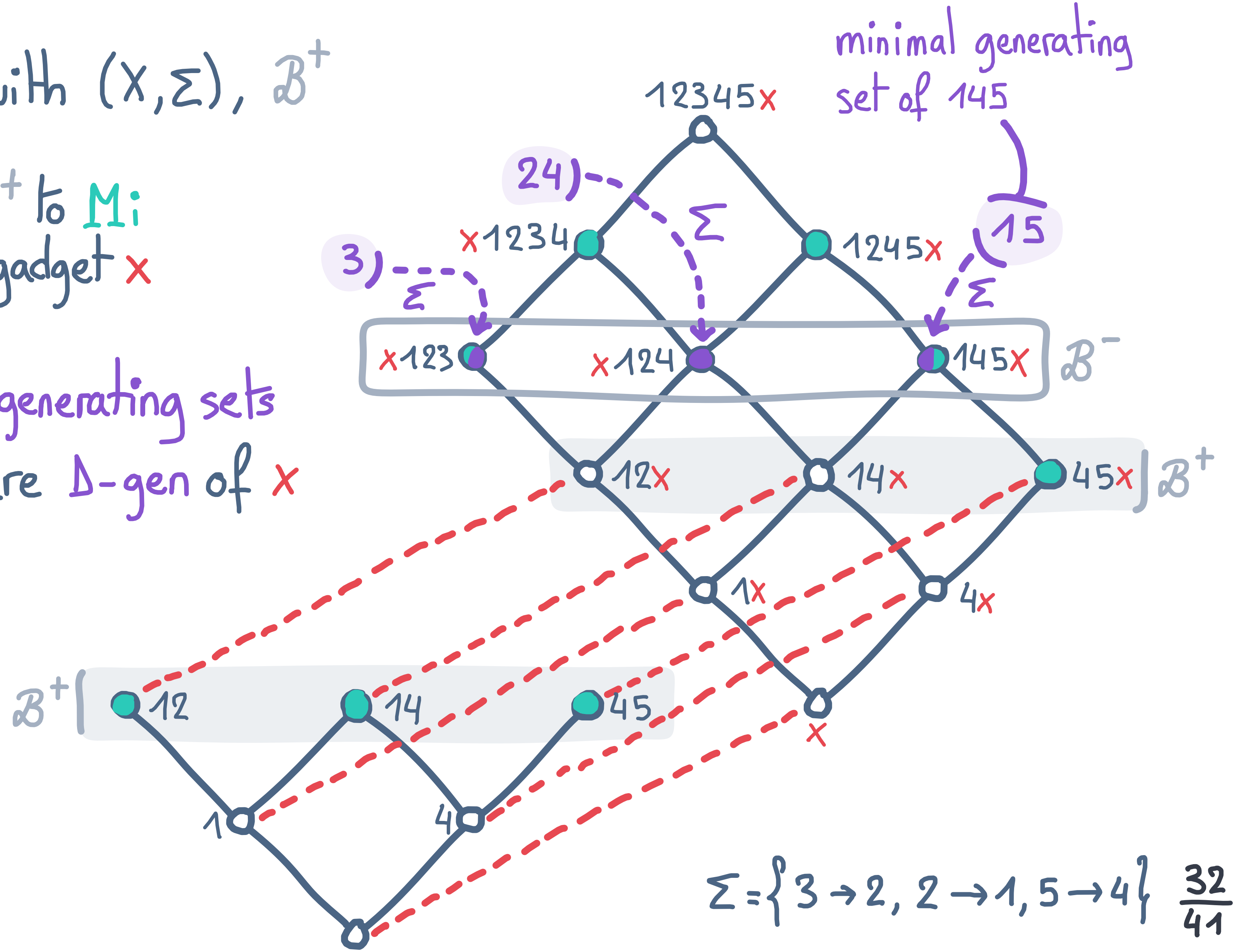
$\Sigma_b = \{3 \to 2,\ 4 \to 2\}$

Intuition: DB-M ≈ Dualization Distr.

① Start with (X, Σ), $\mathcal{B}^+$

② Add $\mathcal{B}^+$ to M_i using gadget x

③ (min.) generating sets of $\mathcal{B}^-$ are Δ-gen of x



$\Sigma = \{3 \to 2, 2 \to 1, 5 \to 4\}$

Long story short

ANV, 23+

DB-M is equivalent to dualization in distributive closure systems

ANV, 23+

DB-M can be solved in output-quasipolynomial time

using Elbassioni, 22

D-base from Σ (DB-IB): given (X, Σ), find (X, Σ_D)

ANV, 23+

DB-IB can be solved with polynomial delay

Our approach: supergraph trasversal

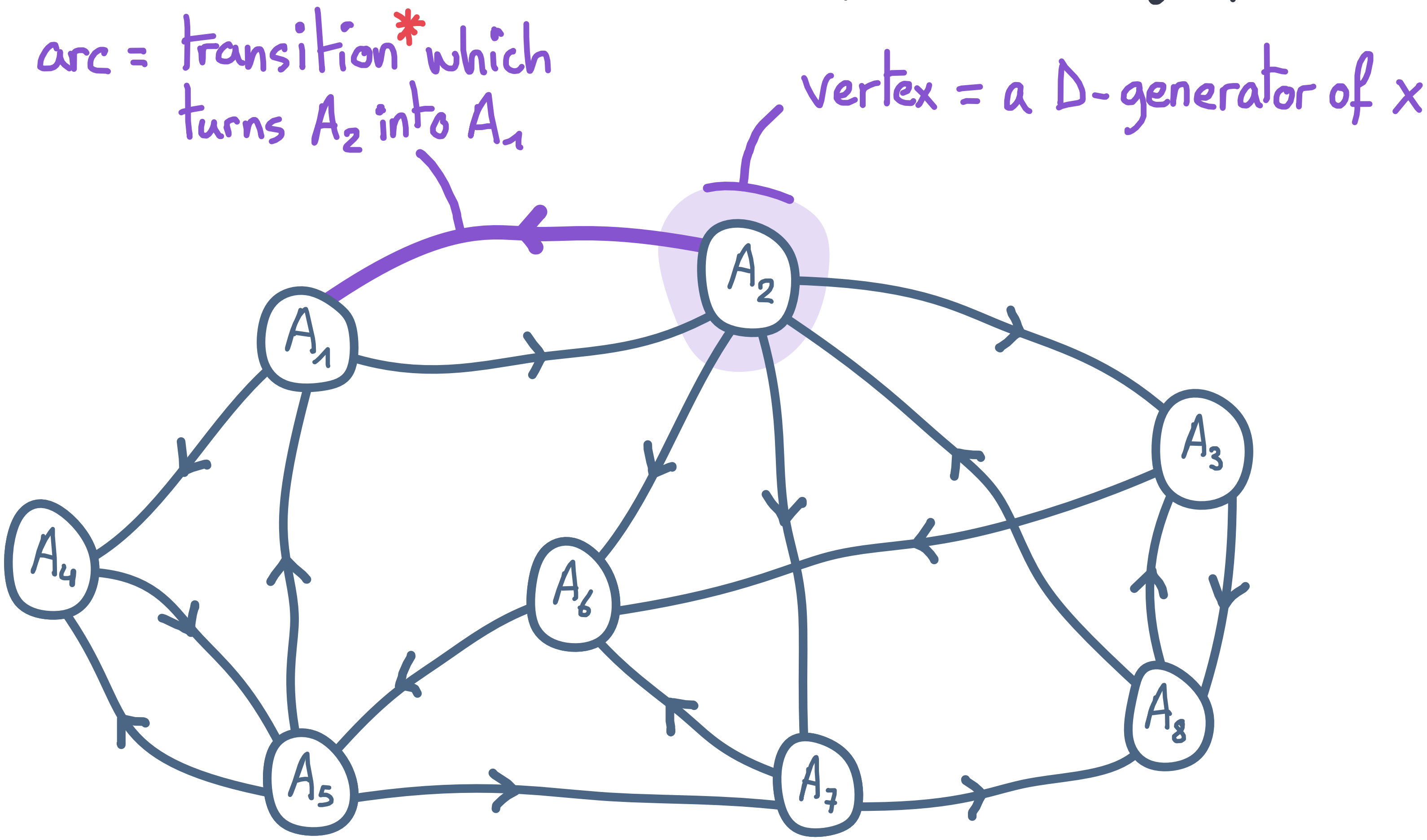
Existing work:

- algorithm using simplification logic Rodriguez et al., 15, 17
  no (output-sensitive) complexity analysis
- poly-delay algorithm listing $\boxed{\Delta\text{-minimal keys}}$ Ennaoui, Nourine, 16
  based on supergraph traversal
  ($\hat{=}$ Δ-gen of some x)

IDEA: use Ennaoui, Nourine, 16 as a blackbox

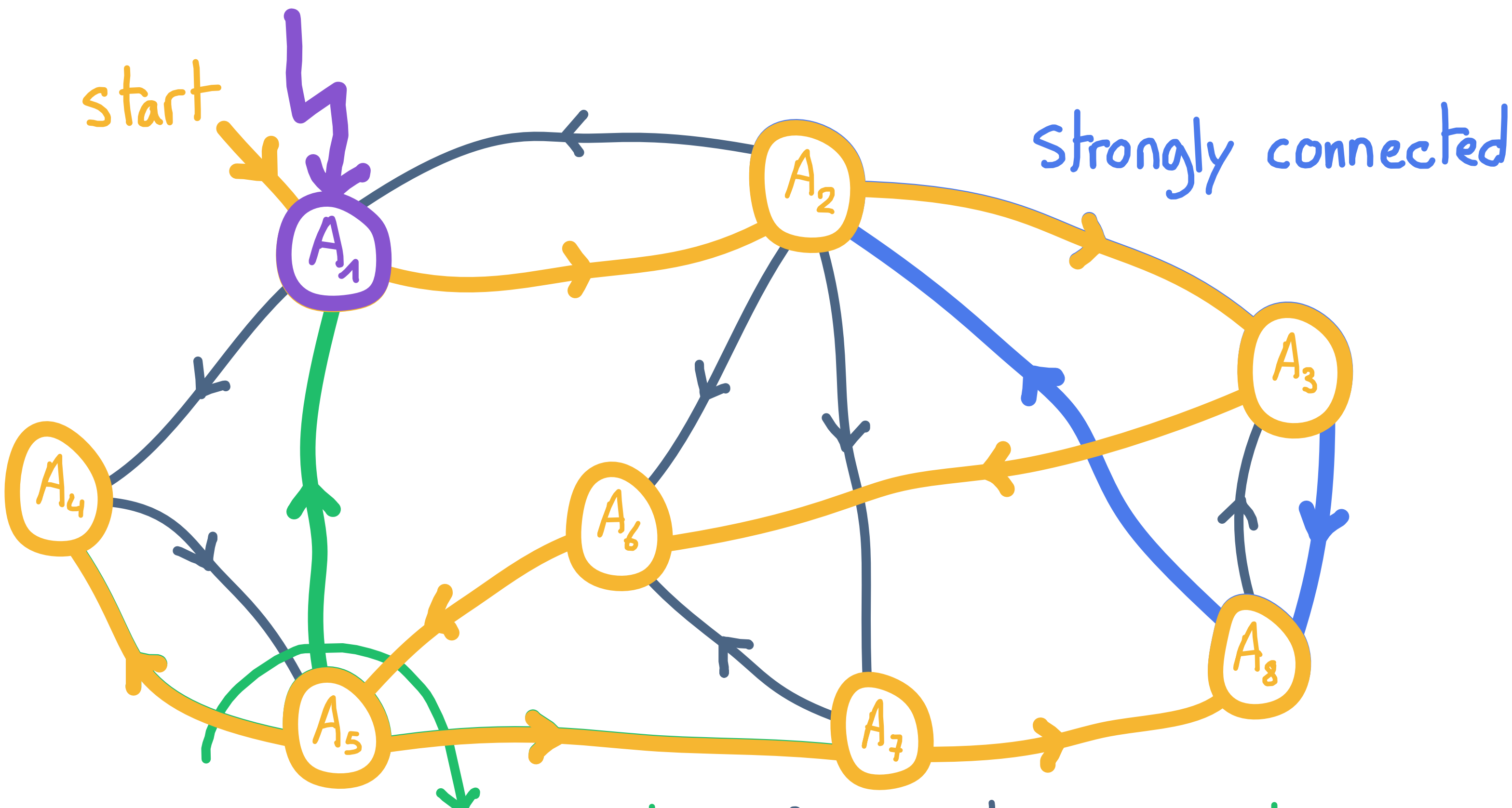
RMK: supergraph traversal also used for minimal Keys
Lucchesi, Osborn, 78 Bérczi et al., 23a

Principle: Supergraph traversal



* transition key idea: substitute $a_2 \in A_2$ with B s.t. $B \to a_2 \in \Sigma$
(greedily) minimize w.r.t. Σ_b

Principle: Supergraph traversal

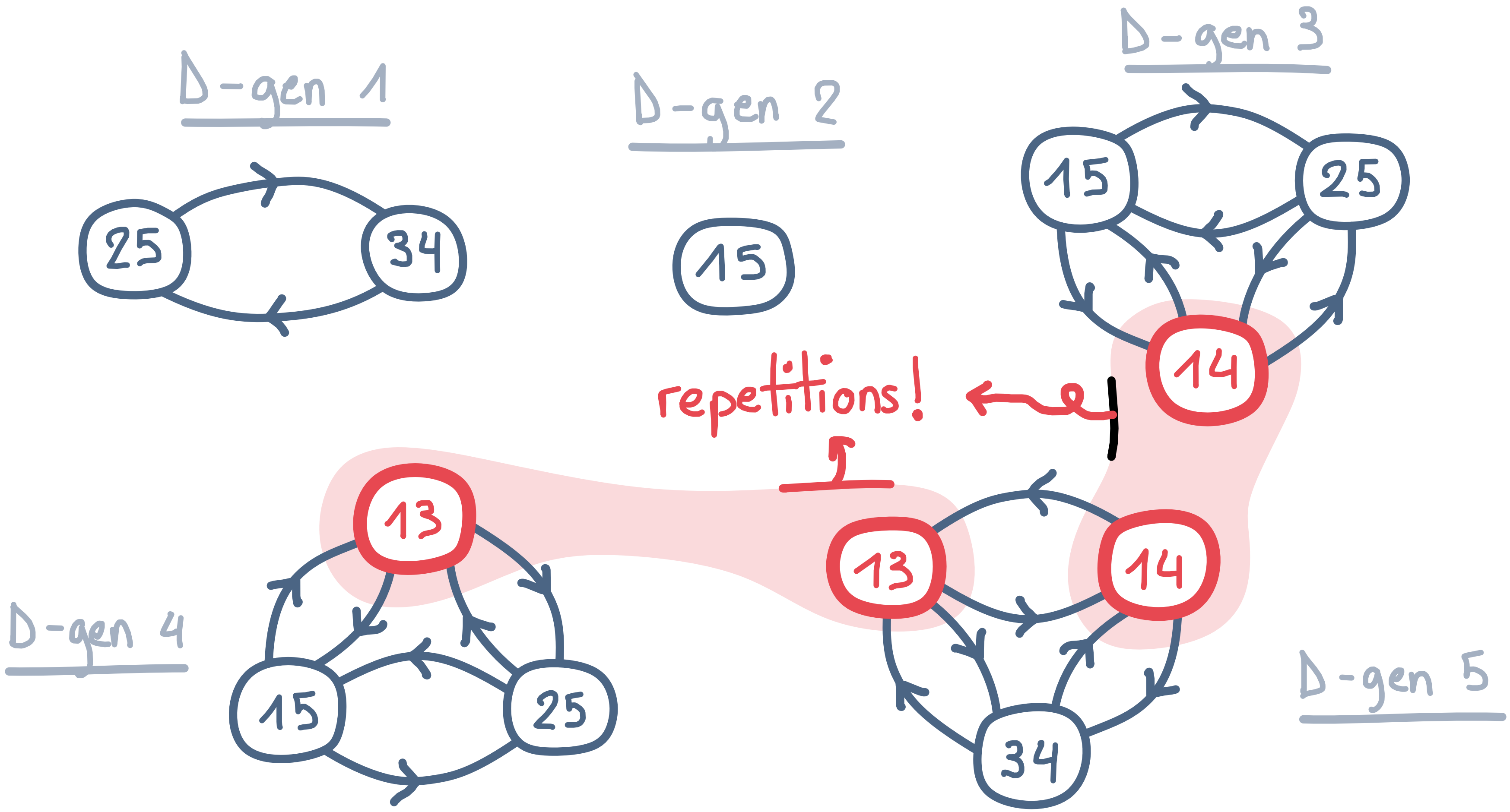


poly number of transitions poly-time computable

1st solution in poly-time + poly transitions + strongly connected

$\Rightarrow$ poly-delay enumeration (with DFS) of D-gen of some x

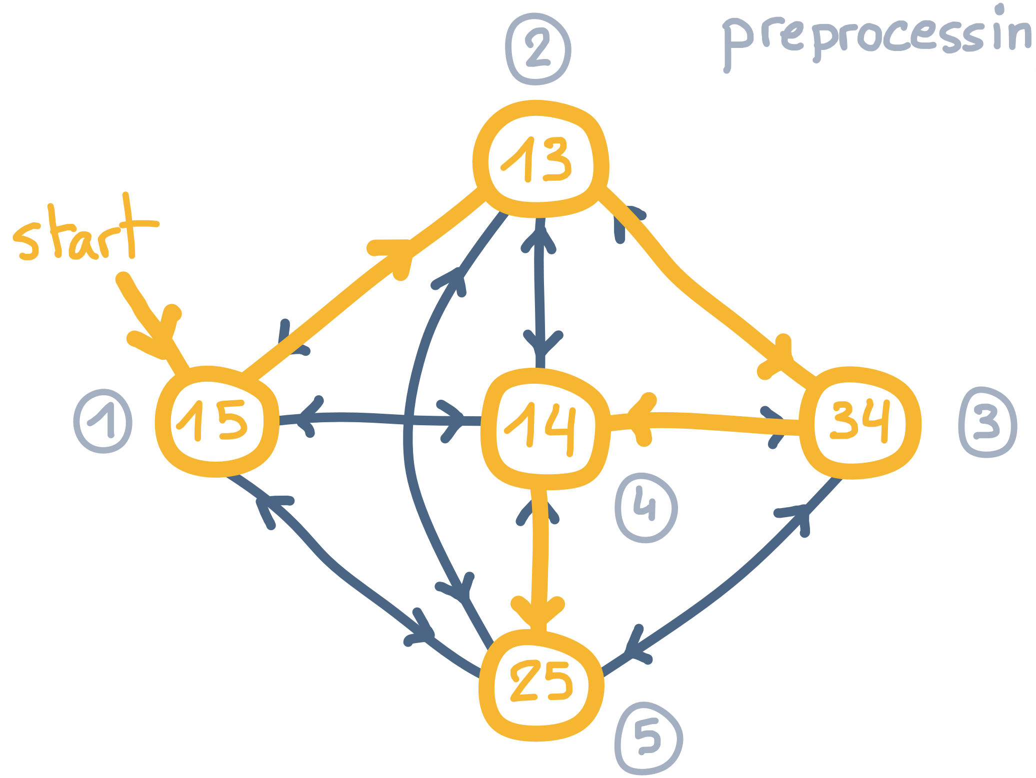
In our case (running ex)



PROB: applying algo on each $x \in X$ yields repetitions

$\Rightarrow$ no guarantee on delay

Fix: merge the graphs



preprocessing

⓪ 3 → 2, 4 → 2

① 15 → 2, 15 → 3, 15 → 4

② 13 → 2, 13 → 5

③ 34 → 5, 34 → 1

④ 14 → 3, 14 → 5

⑤ 25 → 3, 25 → 1, 25 → 4

FIX: take the union of supergraphs

- poly transitions
- 1st solution in poly-time $\forall x \in X$
- strongly connected components

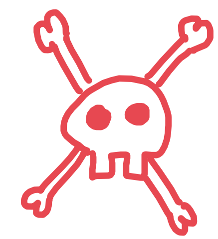
$\Rightarrow$ poly delay enumeration of all D-gens (with DFSs)

Long story short

ANV, 23+

DB-IB can be solved with polynomial delay

with exponential space!



using Ennaoui, Nourine, 16

Finding the D-base:

- output quasi-poly from $\mathcal{M}_i$
- poly-delay from Σ

Other results:

- NP-hardness of finding D-relation (defined from D-base)
- Connection between E-base ($\subseteq$ D-base) and matroids

Further questions:

- Characterize systems with valid E-base
- Similar algorithms for E-base?

Adaricheva, Bernhardt, Liu, Schmidt — Adaricheva et al., 23

Importance of overnight parameters to predict sea breeze on Long Island

2023

Nation, Cabot-Miller, Segal, Lucito, Adaricheva — Nation et al., 21

Combining algorithms to find signatures that predict risk in early stage of stomach cancer

Journal of Computational Biology, 2021

Adaricheva, Nation — Adaricheva, Nation, 17

Discovery of the D-basis in binary table based on hypergraph dualization

Theoretical Computer Science, 2017

Fredman, Khachiyan

Fredman, Khachiyan, 96

On the complexity of dualization of monotone disjunctive normal forms

Journal of Algorithms, 1996

Elbassioni

Elbassioni, 22

On dualization over distributive lattices

Discrete Mathematics and Theoretical Computer Science, 2022

Kavvadias, Sideri, Stavropoulos

Kavvadias et al., 00

Generating maximal models of a Boolean expression

Information Processing Letters, 2000

Rodriguez-Lorenzo, Adaricheva, Cordero, Enciso, Mora Rodriguez et al., 17
Formation of the D-basis from implicational system using Simplification Logic
International Journal on General Systems, 2017

Rodriguez-Lorenzo, Adaricheva, Cordero, Enciso, Mora Rodriguez et al., 15
From an implicational system to its corresponding D-basis
2015

Ennaoui, Nourine Ennaoui, Nourine, 16
Polynomial delay hybrid algorithms to enumerate candidate keys for a relation
BDA, 2016