ISAIM 2024

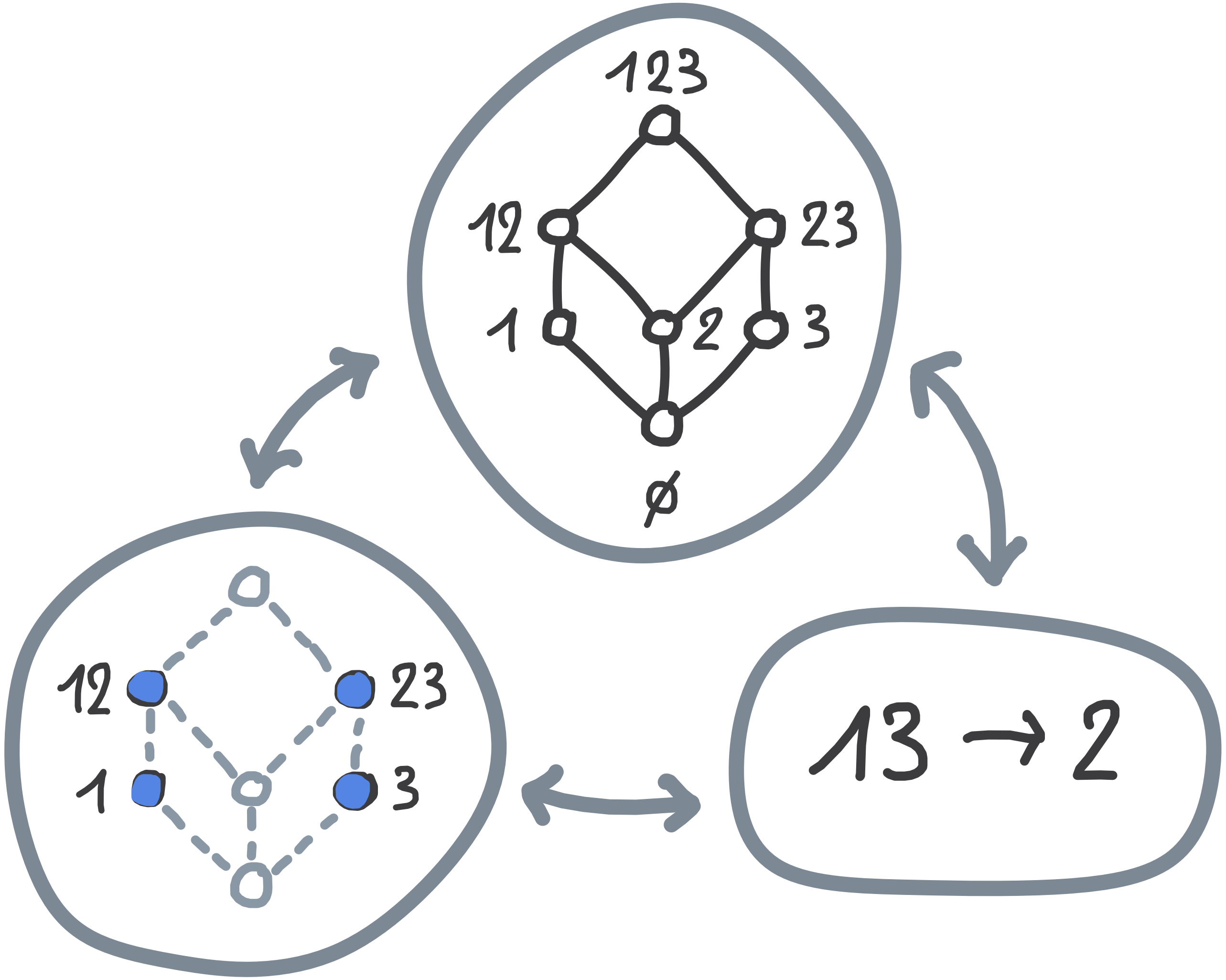
Computing the D-base and D-relation of Finite Closure Systems

Kira Adaricheva — Department of Mathematics, Hofstra University, USA

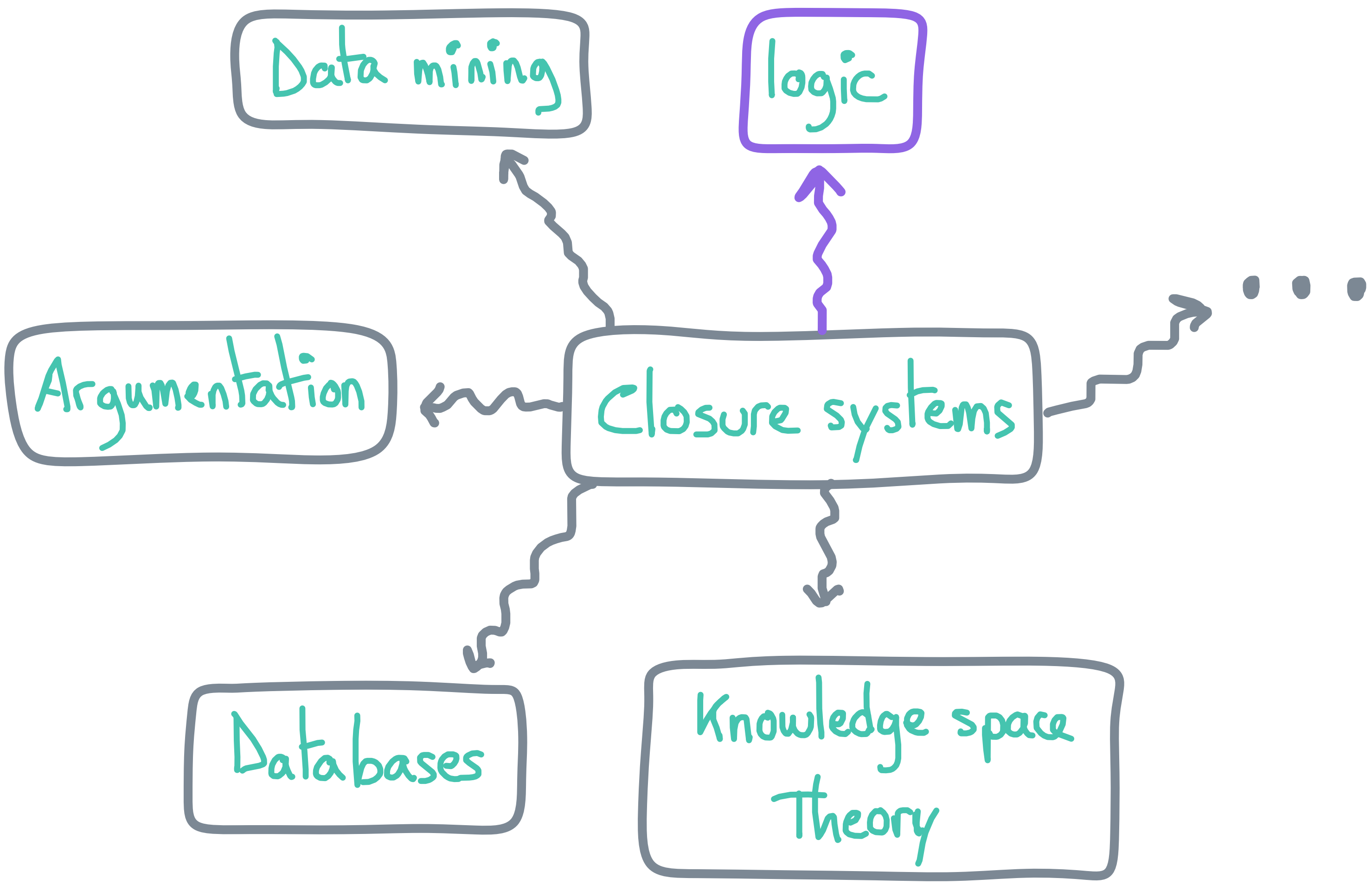
Lhouari Nourine — LIMOS, Université Clermont Auvergne, France

Simon Vilmin — LIS, Aix-Marseille Université, France

Closure systems : what, how, why



What for ?


Data mining
logic
Argumentation
Closure systems
...
Databases
Knowledge space Theory

Did you say Horn functions?

Closure systems	Horn functions
closed set	model
irreducible	characteristic model
implication	(pure) Horn clause
implicational base	(pure) Horn CNF
minimal generator	prime implicate

RMK: hence, every results has its Horn counterpart

Finite closure systems?

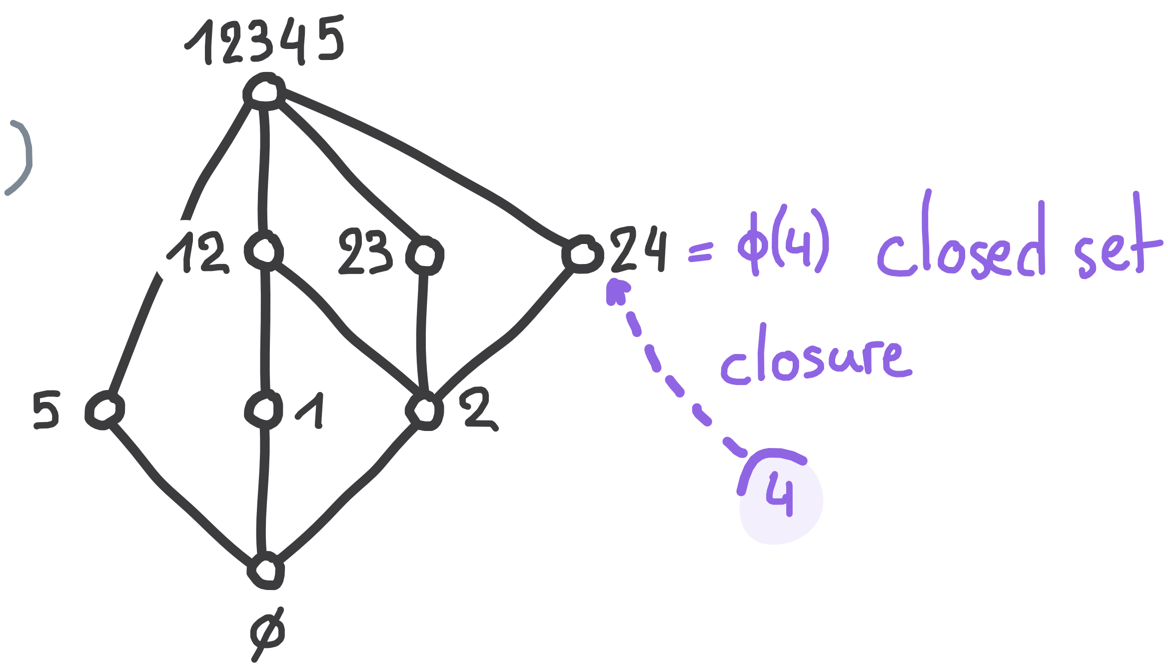
X finite set, $\mathcal{F} \subseteq 2^X$

DEF (closure system): set system $(X, \mathcal{F})$ where

- $X \in \mathcal{F}$
- $F_1, F_2 \in \mathcal{F}$ entails $F_1 \cap F_2 \in \mathcal{F}$ ($\cap$-closed)

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

$X = \{1, 2, 3, 4, 5\}$

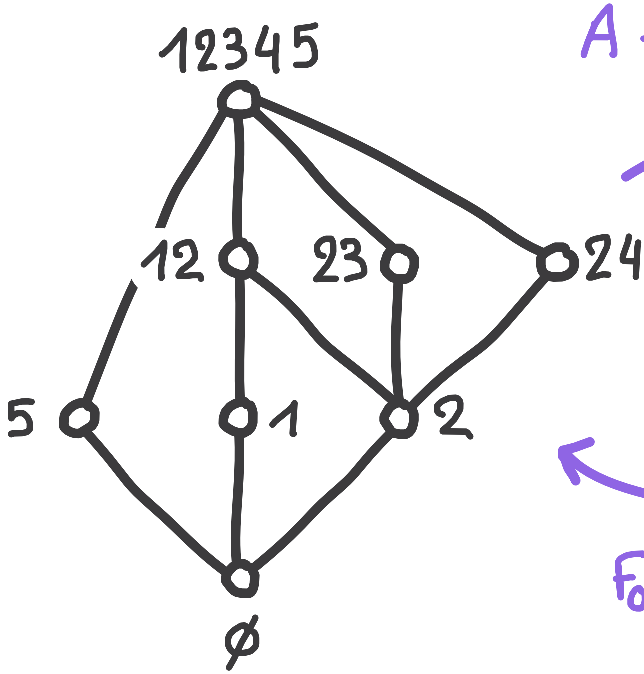


Implications, implicational base (IB)

"If I have A, I have b"

DEF (implicational base, IB):

- **implication** : statement $A \to b$ $(A \subseteq X,\ b \in X)$
- **implicational base** : pair (X, Σ), Σ set of implications



$A \to b \Leftrightarrow b \in \phi(A)$

Forward Chaining

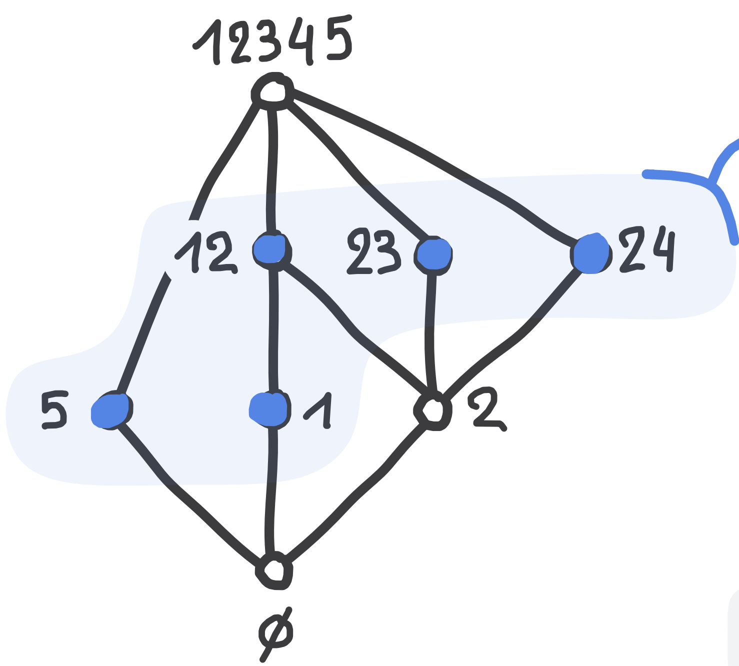
binary implications Σ_b

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

Irreducibles

DEF (irreducible): in $(X, \mathcal{F})$, closed set $M \neq X$ irreducible if not the intersection of other closed sets

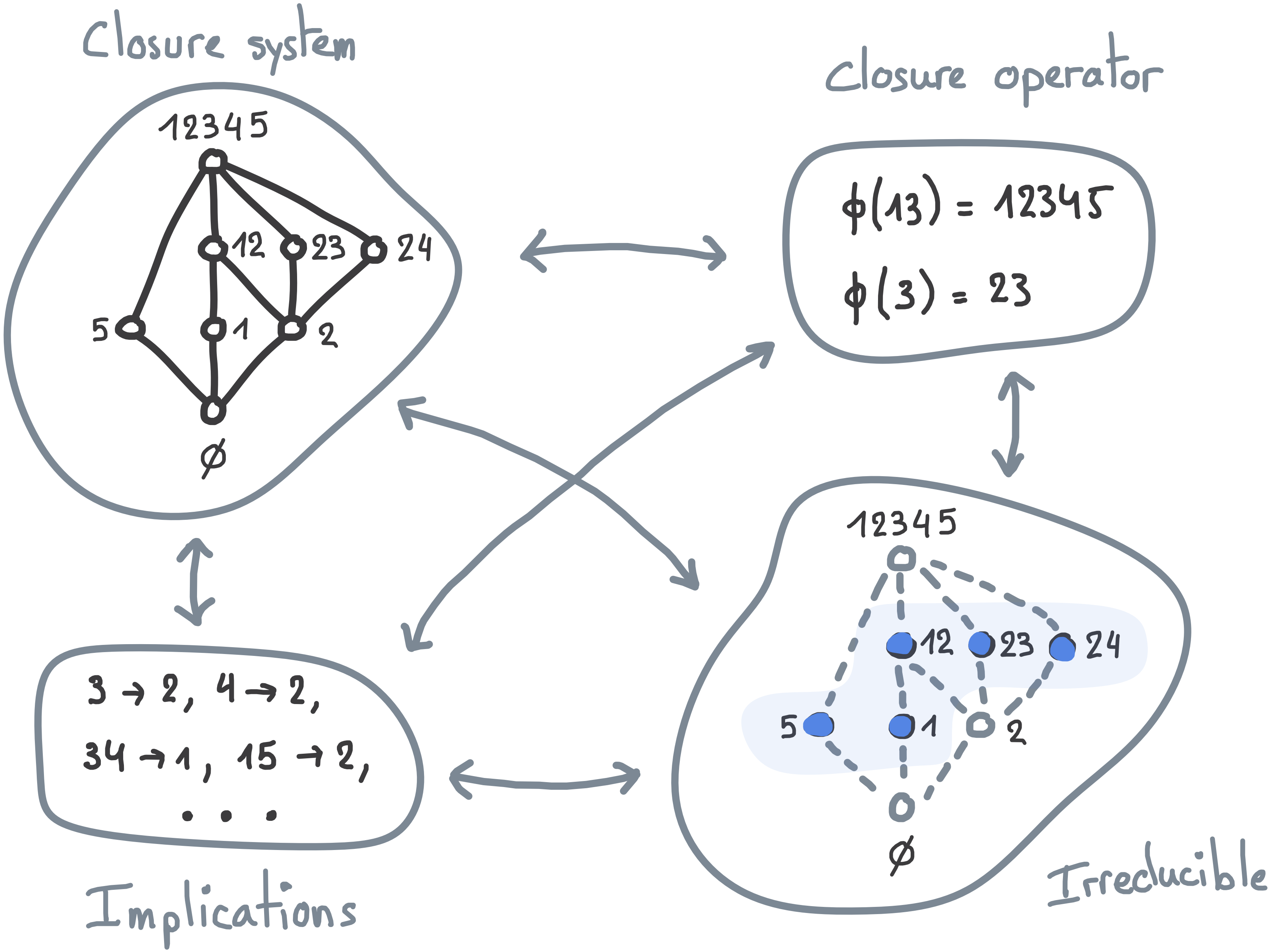
(meet-)



$\mathcal{M}_i$ = irreducibles of $(X, \mathcal{F})$

RMK: $\mathcal{M}_i$ sufficient to recover $\mathcal{F}$

Same information, different POVs



The D-base : our topic of interest

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

Minimal Generators

"What is the minimal way of deriving x?"

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

- prime implicate
- LHS-minimal FD
- circuit of Matroid

binary implications Σ_b

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

$35 \to 2$ ✗ $\quad 5 \to 2$

$234 \to 1$ ✗ $\quad 34 \to 1$

$15 \to 3$ ✓

D-generators, D-base

DEF (D-generator, D-base):

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

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

$$\Sigma_D = \underbrace{\{4 \to 2, 3 \to 2\}}_{\Sigma_b} + \begin{bmatrix} 34 \to 1, & 25 \to 1, & 15 \to 2, \\ 13 \to 2, & 15 \to 3, & 14 \to 3, \\ 25 \to 3, & 15 \to 4, & 25 \to 4, \\ 13 \to 5, & 34 \to 5, & 14 \to 5 \end{bmatrix}$$

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 Δ-base to enjoy its properties

↓ more precisely

Δ-base from Mi (ΔB-M): given Mi, find (X, Σ_Δ)

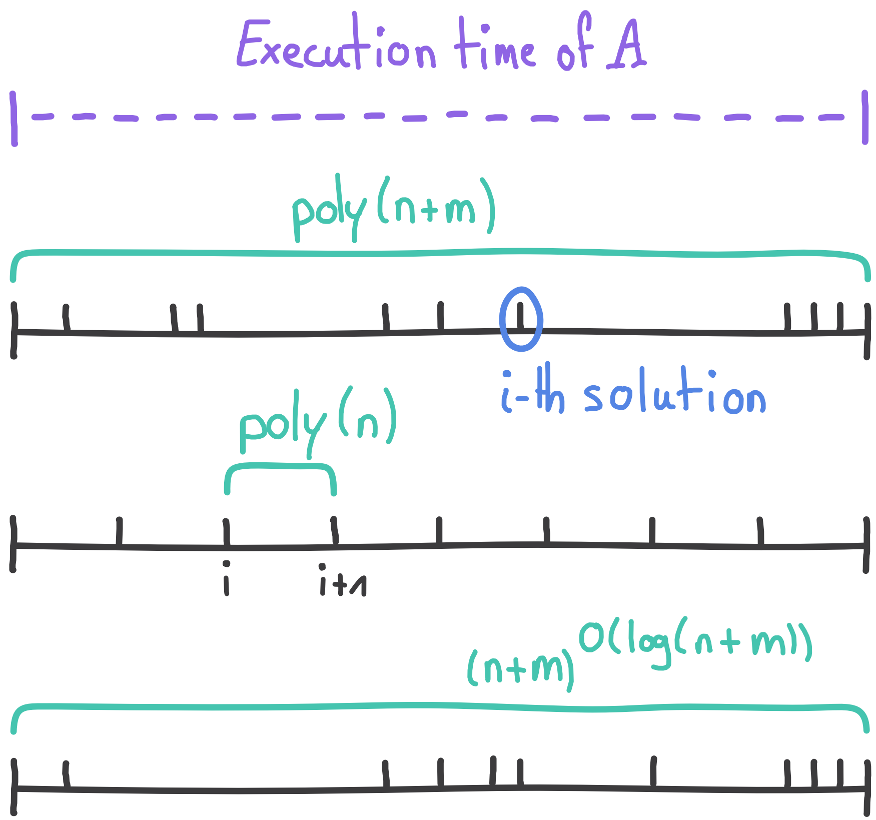
Δ-base from Σ (ΔB-IB): given (X, Σ), find (X, Σ_Δ)

Enumeration: output-sensitive complexity

Each of size poly(x)

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

Enumeration algorithm $\mathcal{A}$
x of size n, R(x) of size m



Output polynomial time

polynomial delay

Output quasi-polynomial time

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

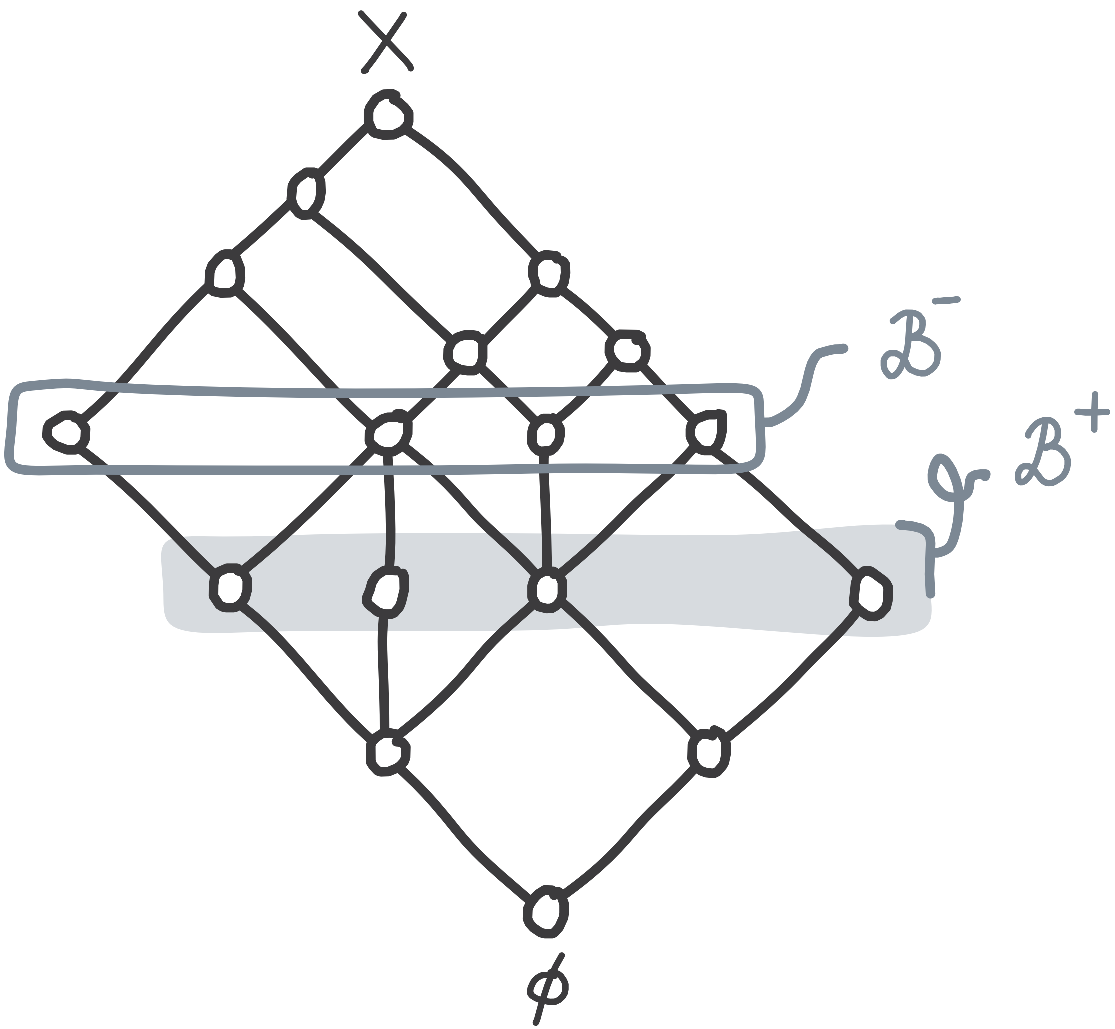
Our approach: dualization

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
⇒ use dualization in distributive closure systems

Dualization (with Σ)

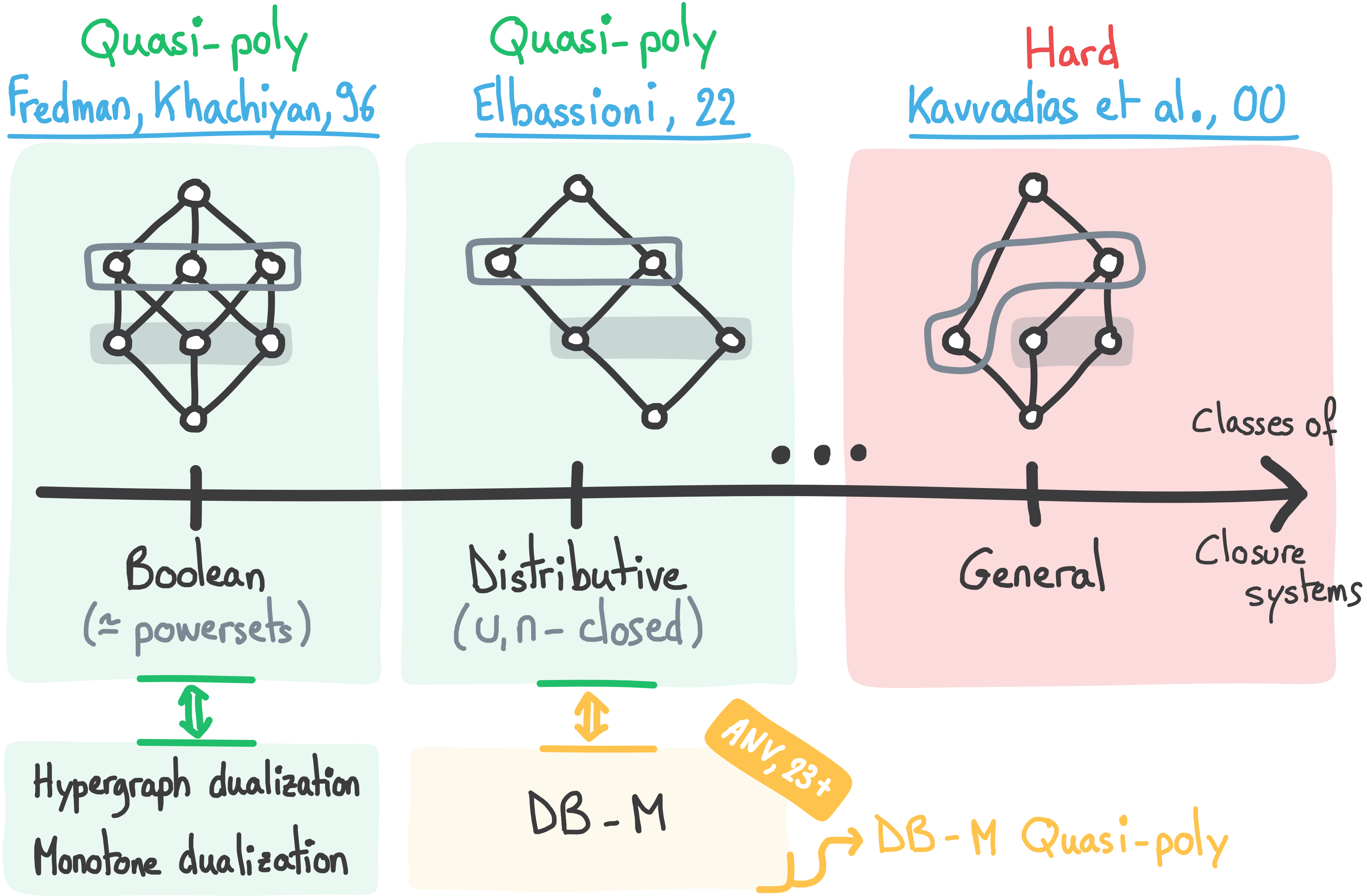


$\mathcal{B}^-$ and $\mathcal{B}^+$ are dual:

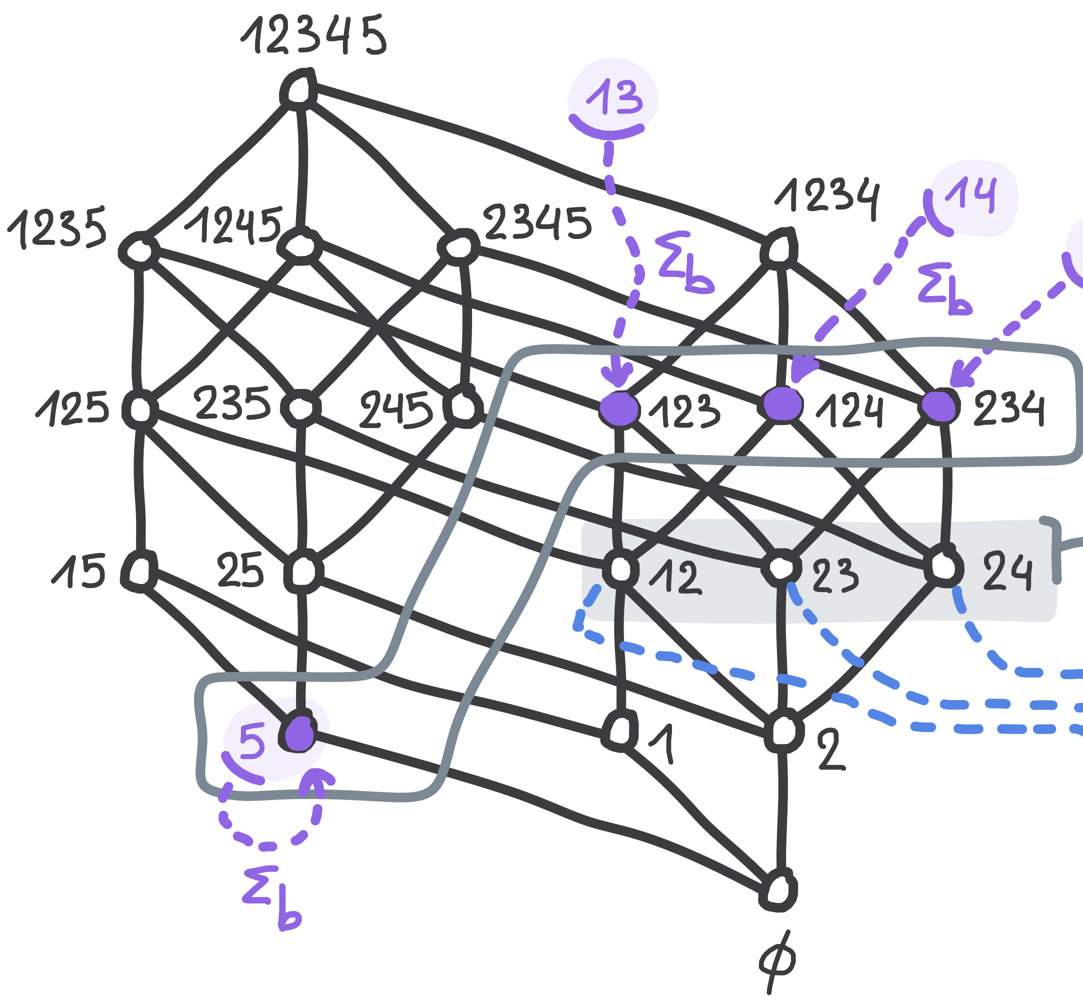
- $\downarrow \mathcal{B}^+ \cup \uparrow \mathcal{B}^- = \mathcal{F}$
- $\downarrow \mathcal{B}^+ \cap \uparrow \mathcal{B}^- = \emptyset$

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

Dualization complexity (with Σ) and DB-M

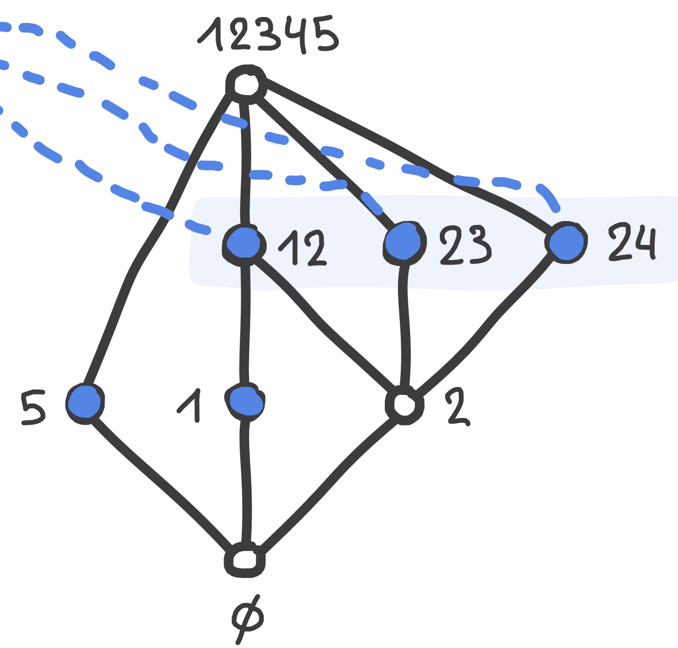


Intuition: DB-M ≈ Dualization Distr.



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

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



① distributive closure system of

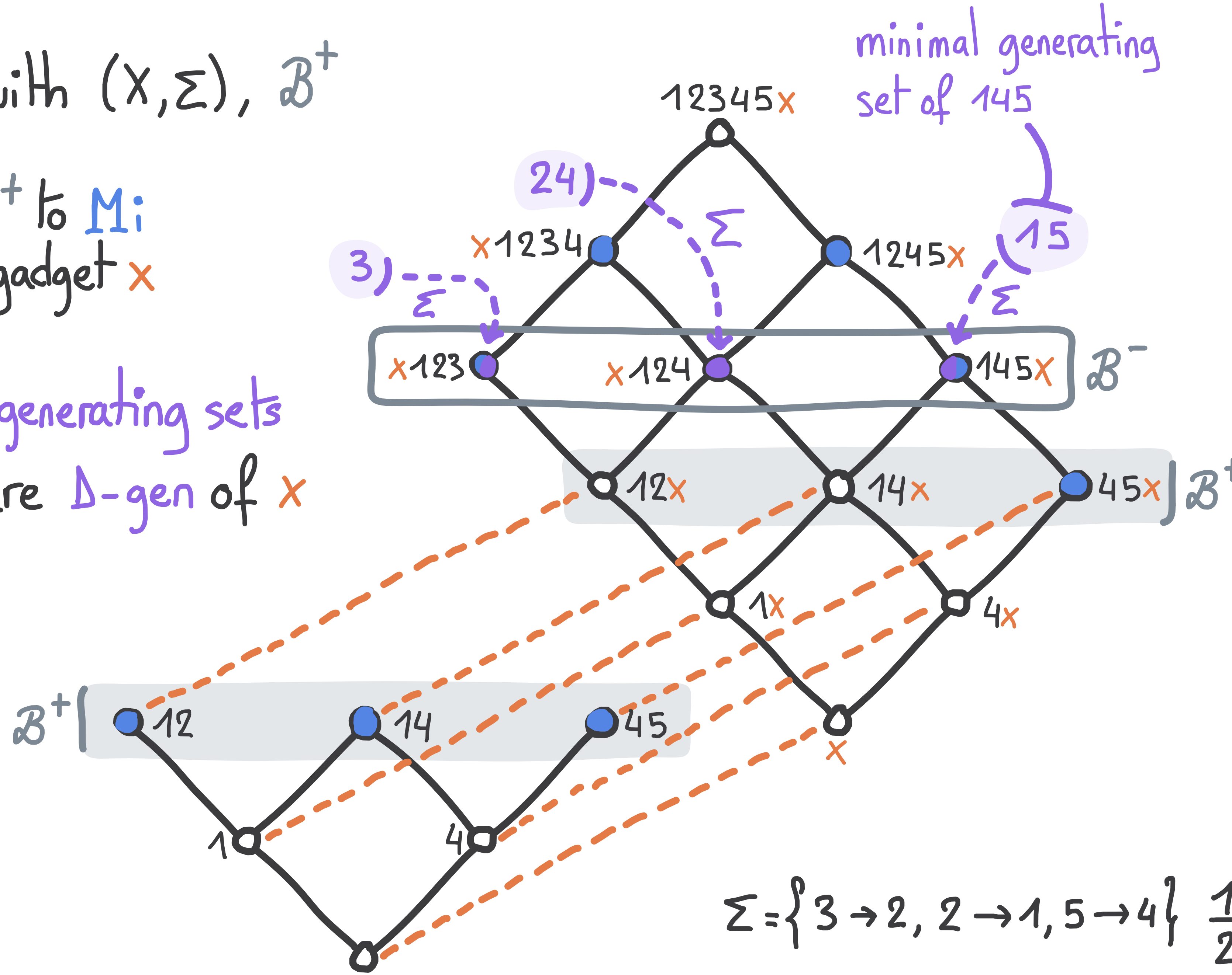
$\Sigma_b = \{3 \to 2,\ 4 \to 2\}$

Intuition: DB-M ≈ Dualization Distr.

① Start with (X, Σ), $\mathcal{B}^+$

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

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



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

$\frac{16}{24}$

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

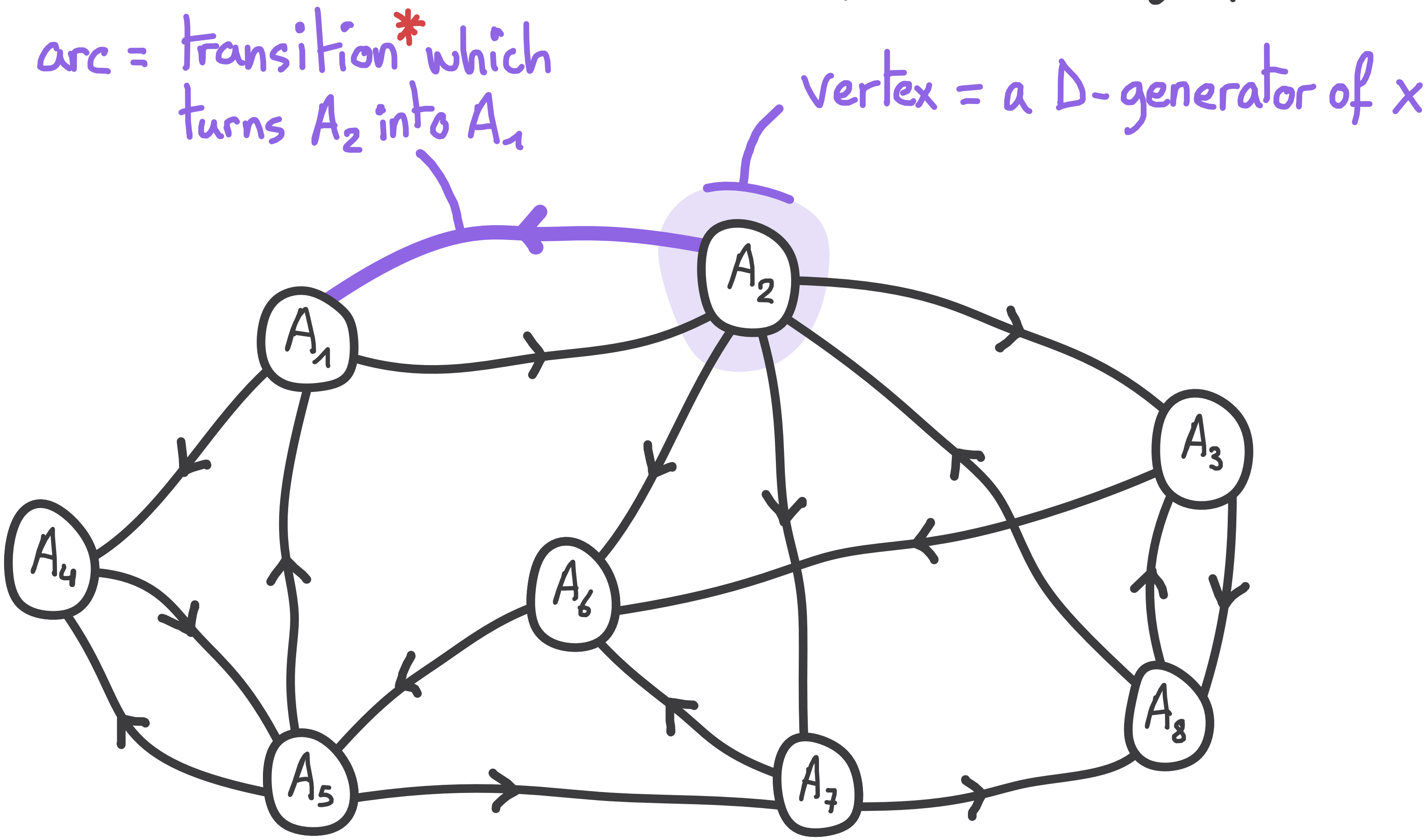
Existing work:

- algorithm using simplification logic <u>Rodriguez et al., 15, 17</u>
  no (output-sensitive) complexity analysis
- poly-delay algorithm listing Δ-minimal keys <u>Ennaoui, Nourine, 16</u>
  based on supergraph traversal
  (≙ Δ-gen of some x)

IDEA: use <u>Ennaoui, Nourine, 16</u> as a blackbox

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

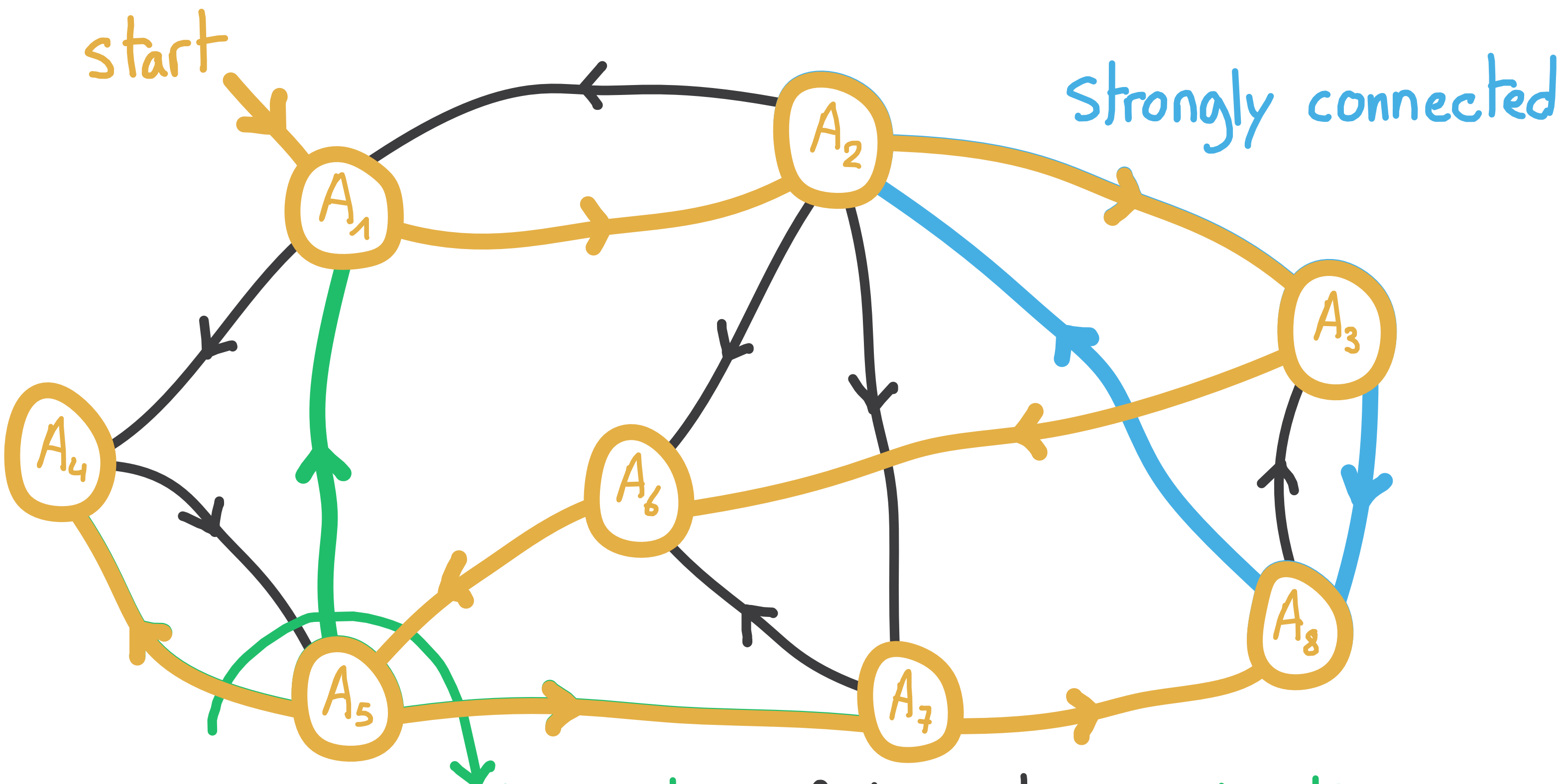
Principle: Supergraph traversal



* 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

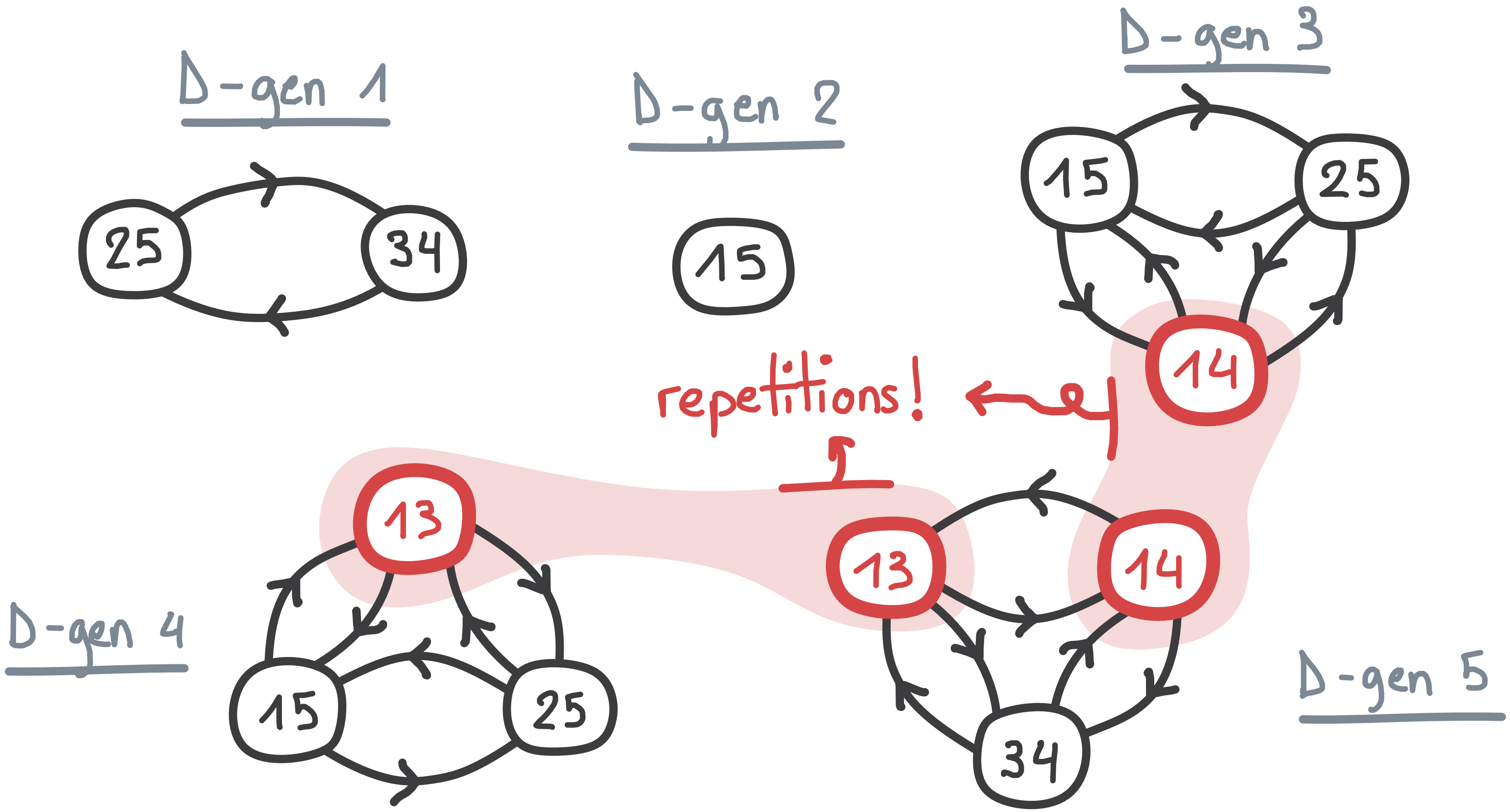


poly number of transitions poly-time computable

poly transitions + strongly connected

⇒ poly-delay enumeration (with DFS) of D-gen of some x

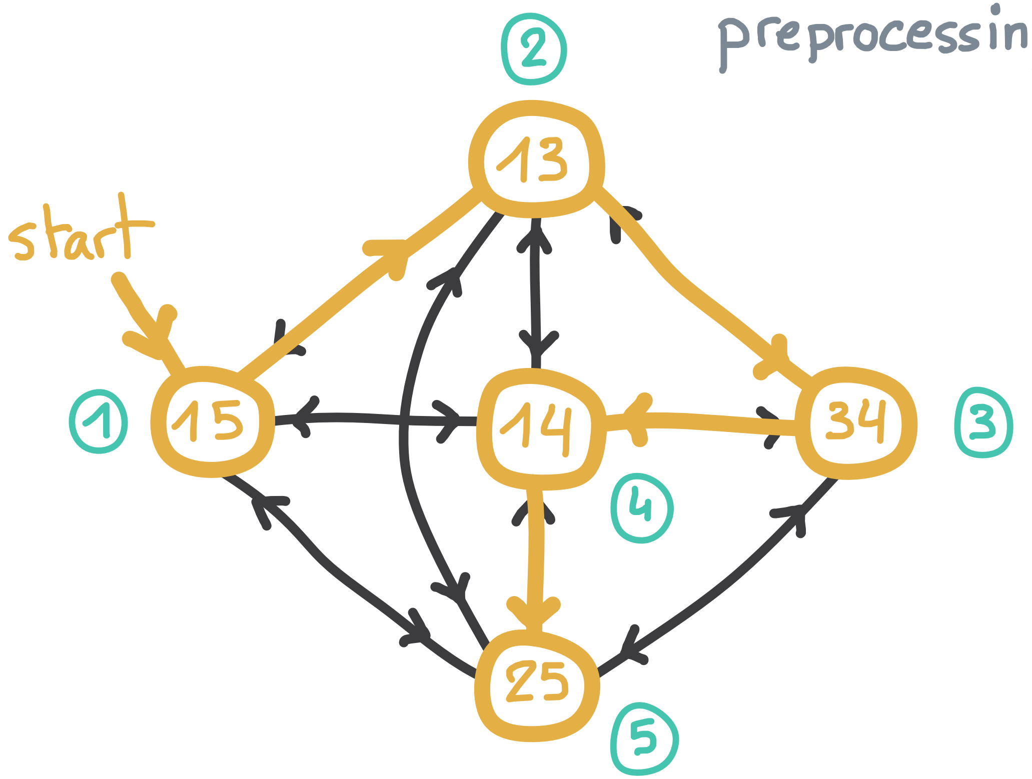
In our case (running ex)



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
- strongly connected components

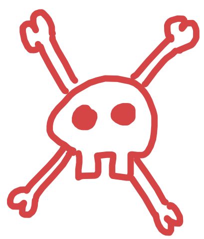
⇒ poly delay enumeration of all D-gens (with DFSs)

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