Closure Systems and Their representations

Simon Vilmin

March, 23, 2023

MALOTEC Seminar

Back to school

- Knowledge Space Theory [Doignon, Falmagne, 1985]:

  "Automatically assess the knowledge of students"

- Some questions of an automated test
  1. Graphically solve $4x^2 - 3x + 2 = 0$ → graphical resolution
  2. Figure out $\frac{(\sqrt{4} \times \sqrt{9})}{3} - \frac{6 \times 7}{\sqrt{144}}$ → arithmetic
  3. Find the discriminant of $3x^2 - x + 8$ → formula of discriminant
  4. Study the polynomial $7x^2 + 11x - 5$ → study of 2nd order polynomial

Each question corresponds to a problem or item

What is your score?

- Some students took the test!

	1	2	3	4
Wolf	X			
Lil		X	X	
Lazuli		X	X	X
Folavril	X	X		X
Dupont		X		

- Lil masters the items 2 and 3 but not the items 1 and 4

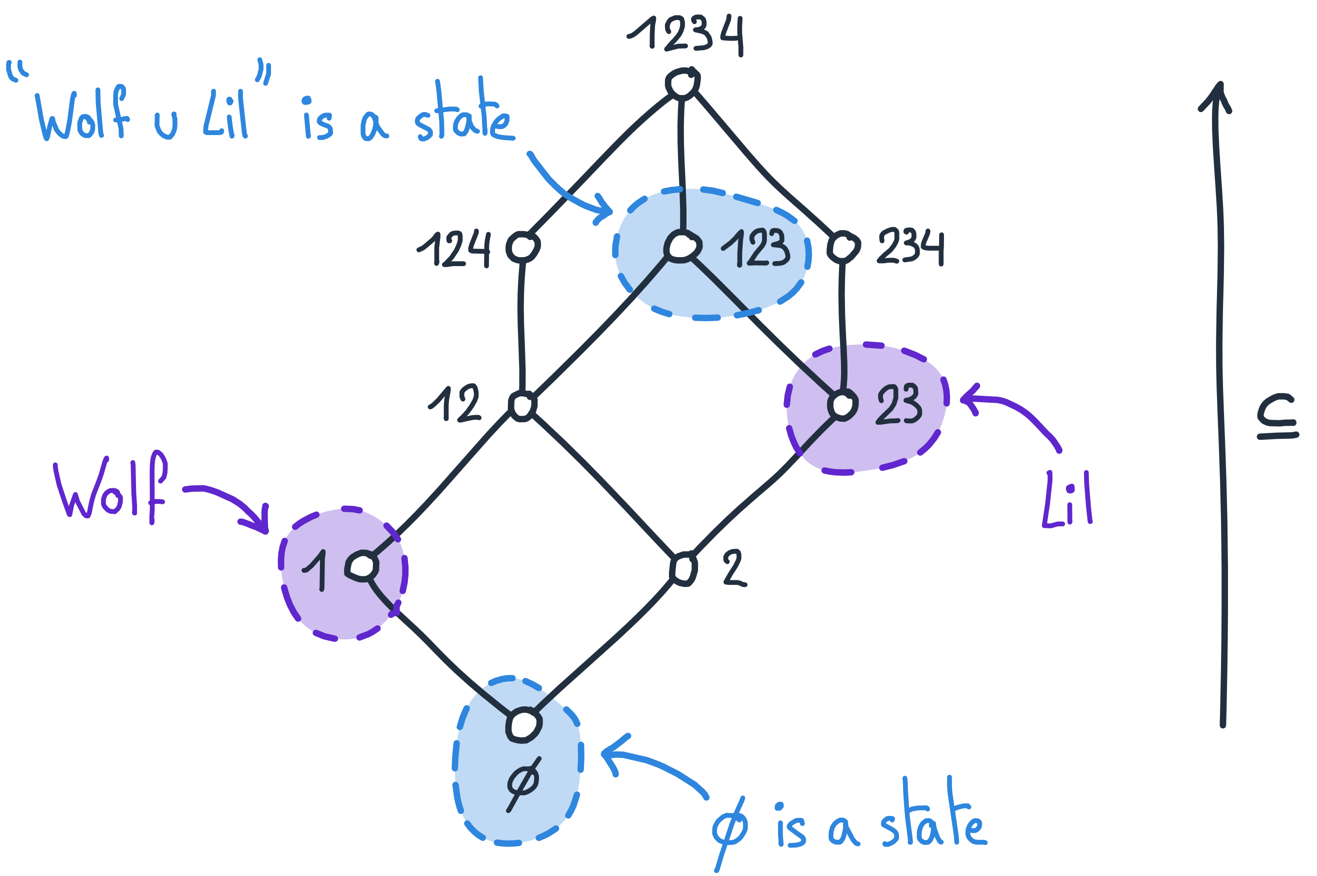
- $\boxed{23}$ is the (knowledge) state of Lil

Abbreviation of $\{2,3\}$

Knowledge spaces

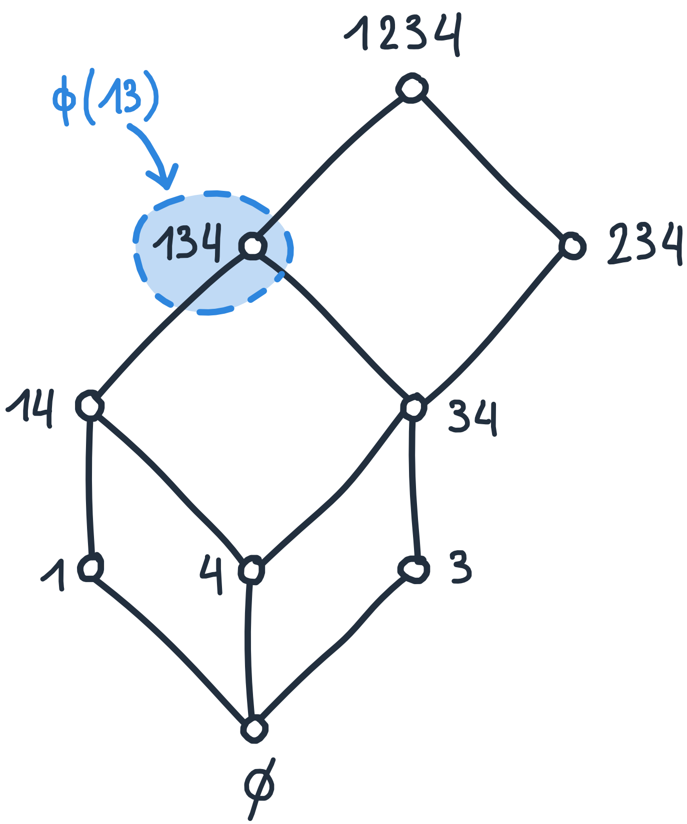
- Knowledge space $(X, \mathcal{K})$: set X of items, collection $\mathcal{K}$ of states over X s.t.
  - $\emptyset \in \mathcal{K}$
  - $K_1, K_2 \in \mathcal{K}$ implies $K_1 \cup K_2 \in \mathcal{K}$

Mathematical but reasonable assumptions



In fact, closure systems

DEF. A closure system is a pair $(X, \mathscr{C})$ where X is a set, the groundset, and $\mathscr{C}$ a collection of subsets of X satisfying $X \in \mathscr{C}$ and $C_1 \cap C_2 \in \mathscr{C}$ for every $C_1, C_2 \in \mathscr{C}$.



- Sets in $\mathscr{C}$ are closed sets
- the pair $(\mathscr{C}, \subseteq)$ is a lattice
- Induces a closure operator ϕ:
  - $\phi(A)$ = minimal closed set including A

$\boxed{\text{Closure system = complement of a Knowledge space}}$

The problem with closure systems

- Closure systems are ubiquitous
  - Knowledge Space Theory, Argumentation theory, Propositional logic, Formal Concept Analysis (FCA), Databases, ...

In KST, you cannot ask teachers for a set of states

- But they have HUGE size and can be hard to understand

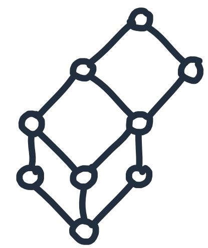
If $|X| = n$, $\mathcal{C}$ can have up to 2^n closed sets

We need implicit representations!

meet-irreducible elements

Implicational Bases

Outline



closure systems

Same knowledge, different representations:
What is the complexity of translating between them ?

meet-irreducible elements

implicational bases

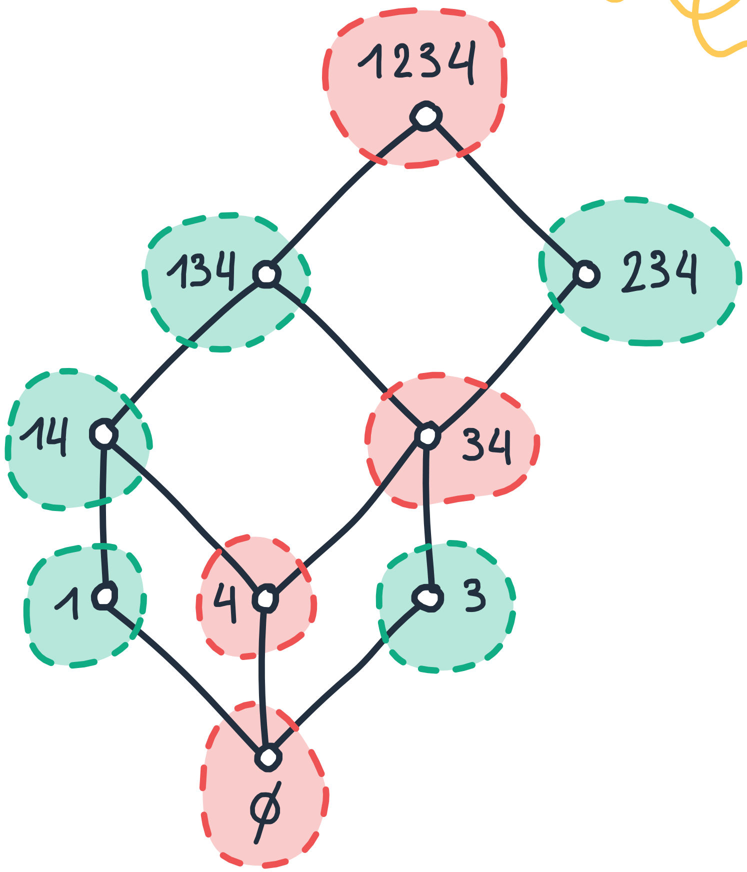
Step 1. What is this ?

Meet-irreducible elements : intuition

- We want to compactly represent a closure system $(X, \mathcal{C})$

IDEA : find a small subset of $\mathcal{C}$ which conveys all the information of $(X, \mathcal{C})$

How?



- Use properties of closure systems
  - $X \in \mathcal{C}$ trivially holds $\rightarrow$ useless
  - $C \in \mathcal{C}$ is obtained by intersections $\rightarrow$ useless
  - $C \in \mathcal{C}$ is not obtained by intersections $\rightarrow$ C is crucial to $(X, \mathcal{C})$, it is irreducible

The irreducible closed sets form the minimal amount of sets needed to rebuild $\mathcal{C}$

Meet-irreducible elements: definition

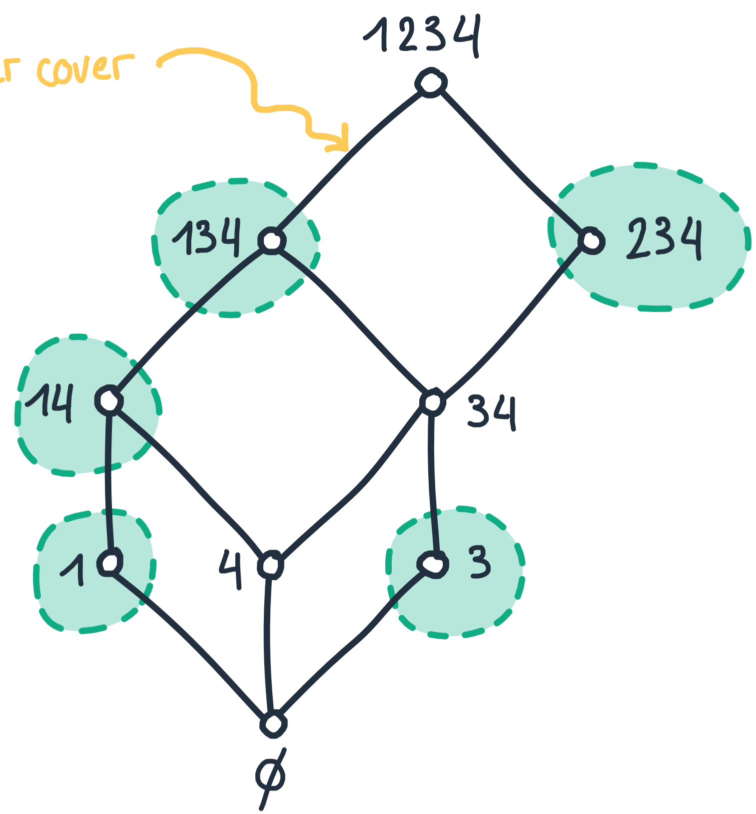
Def. Let $(X, \mathcal{C})$ be a closure system, and let $M \in \mathcal{C}$. The closed set M is **meet-irreducible** if $M \neq X$ and for every $C_1, C_2 \in \mathcal{C}$,

$$M = C_1 \cap C_2 \text{ implies either } M = C_1 \text{ or } M = C_2.$$

We denote by $\mathcal{M}i(\mathcal{C})$ the set of meet-irreducible elements of $(X, \mathcal{C})$.

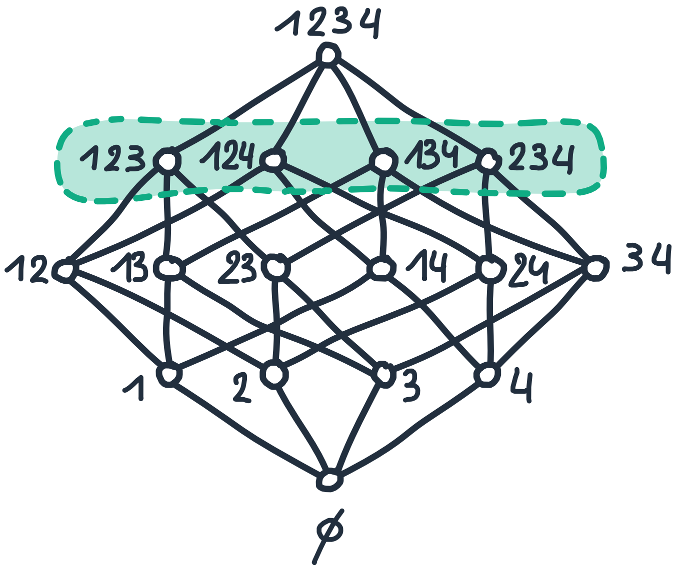
Meet-irreducible $\leftrightarrow$ unique upper cover

$(\mathcal{C}, \subseteq)$



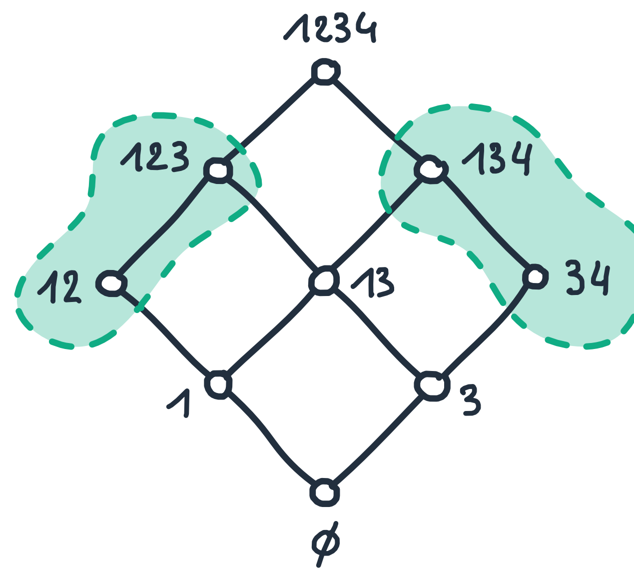
$\mathcal{M}i(\mathcal{C}) = \{3, 234, 1, 14, 134\}$

Meet-irreducible elements : more examples



- $(X, \mathcal{C})$ Boolean cube, $|X| = n$
- $Mi(\mathcal{C}) = \{ X \setminus \{x\} \mid x \in X \}$
- $\boxed{2^n = |\mathcal{C}| \gg |Mi(\mathcal{C})| = n}$

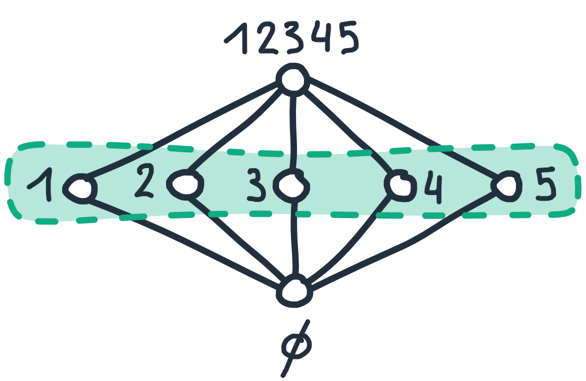
exponential gap



K is fixed

- $(X, \mathcal{C})$ is a $\boxed{\text{k-dim}}$ grid, $|X| = n$
- $Mi(\mathcal{C}) = \{ X \setminus \phi(x) \mid x \in X \}$
- $\boxed{n^k \approx |\mathcal{C}| > |Mi(\mathcal{C})| = n}$

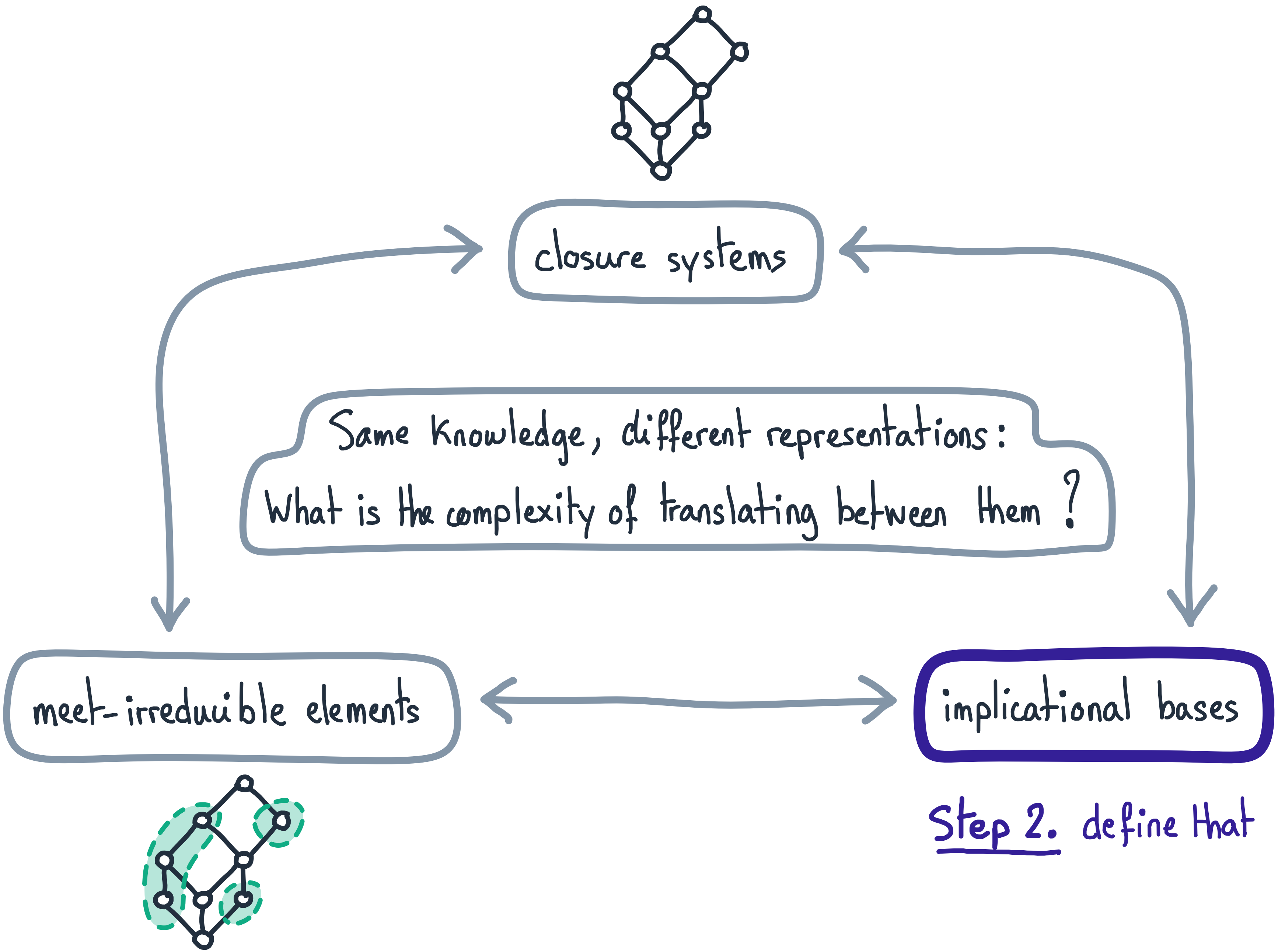
polynomial gap



- $(X, \mathcal{C})$ is a diamond, $|X| = n$
- $Mi(\mathcal{C}) = \{ \{x\} \mid x \in X \}$
- $\boxed{n+2 = |\mathcal{C}| \approx |Mi(\mathcal{C})| = n}$

constant gap

Outline

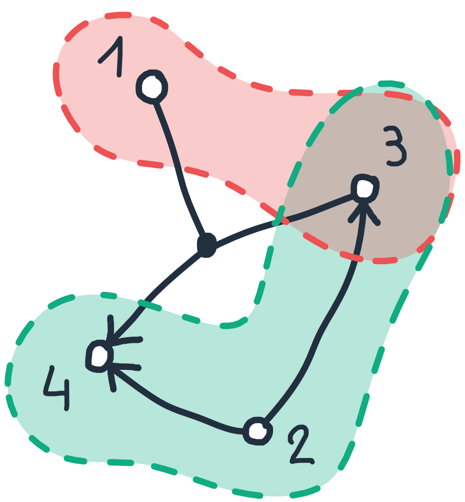

closure systems
Same knowledge, different representations:
What is the complexity of translating between them?
meet-irreducible elements
implicational bases
Step 2. define that

Implications: syntax and semantic

Def. Let X be a set. An implication over X is an expression $A \to B$ where $A, B \subseteq X$. An implicational base is a pair (X, Σ) where Σ is a set of implications over X.

premise: A — conclusion: B

Def. Let $A \to B$ be an implication over X, and $C \subseteq X$. Then, C satisfies $A \to B$ if $A \subseteq C$ implies $B \subseteq C$. If (X, Σ) is an implicational base, C satisfies Σ if it satisfies each implication of Σ.

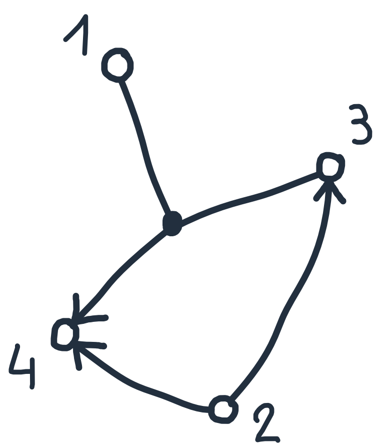


- $X = \{1, 2, 3, 4\}$, $\Sigma = \{13 \to 4,\ 2 \to 34\}$
- 13 does not satisfy Σ $(13 \to 4)$
- 234 satisfies Σ

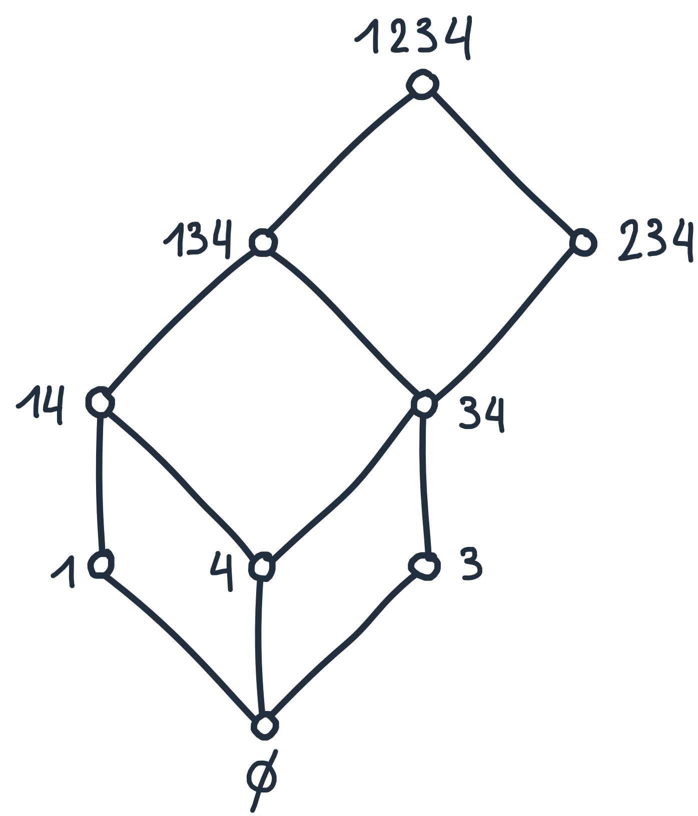
Implications and closure systems

- Define $\mathcal{C}(\Sigma) = \{C \subseteq X \mid C \text{ satisfies } \Sigma\}$, what do we have ?
  - $X \in \mathcal{C}(\Sigma)$ holds
  - if $C_1, C_2 \in \mathcal{C}(\Sigma)$, necessarily $C_1 \cap C_2 \in \mathcal{C}(\Sigma)$ also holds

An implicational base (X, Σ) represents a closure system $(X, \mathcal{C}(\Sigma))$



$\Sigma = \{13 \to 4,\ 2 \to 34\}$



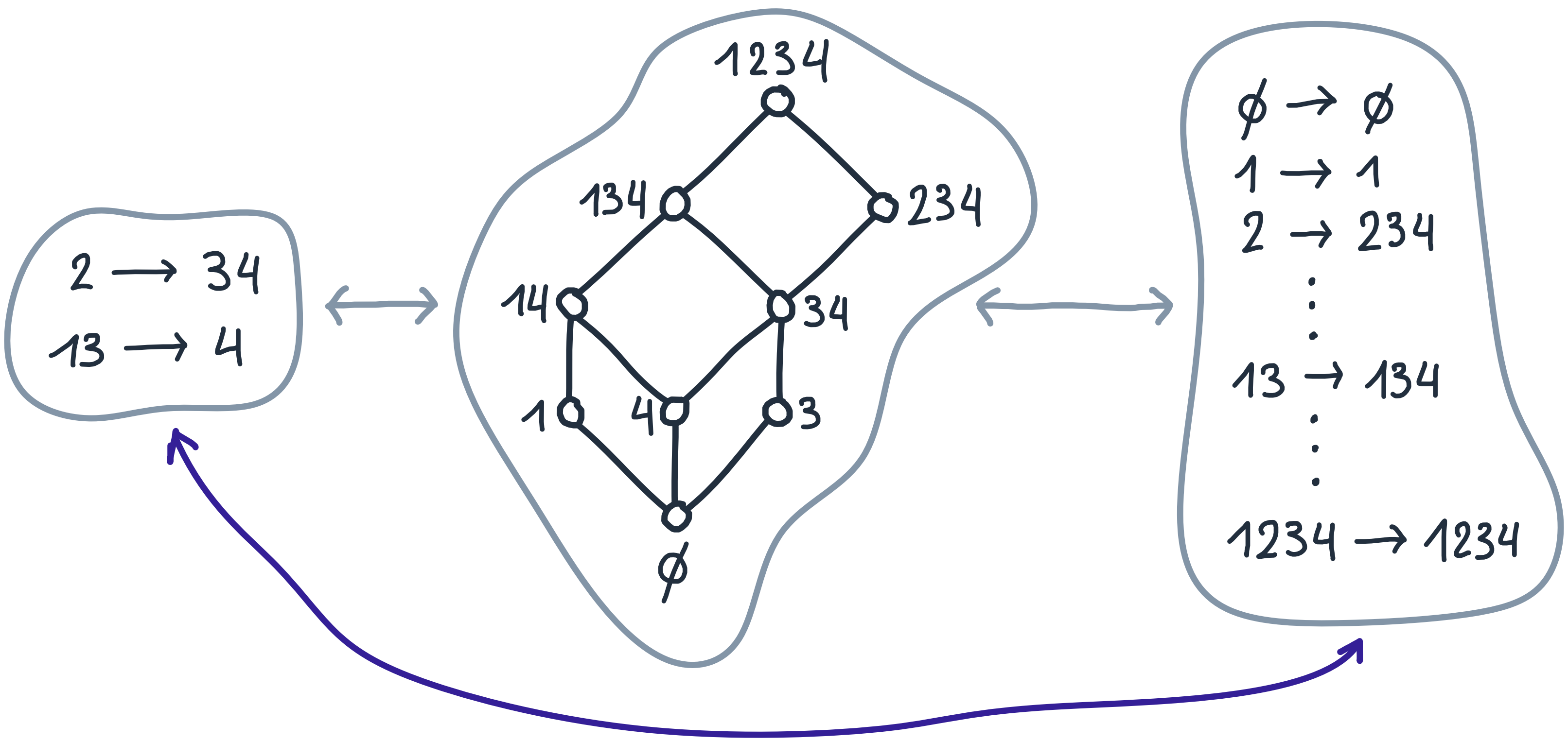
Closure systems and implications

$$\begin{tikzcd} \boxed{\text{implicational bases}} \arrow[r, bend left, "\text{encode}"] & \boxed{\text{closure systems}} \arrow[l, bend left, "\text{encoded by ? YES!}"] \end{tikzcd}$$

- Given $(X, \mathcal{C})$, any closed set including A also includes $\phi(A)$ — Minimal closed set including A
  - The closed sets satisfy $A \to \phi(A)$
- Put $\Sigma = \{A \to \phi(A) \mid A \subseteq X\}$. If $A \notin \mathcal{C}$, A does not satisfy Σ
  - (X, Σ) represents $(X, \mathcal{C})$

<u>THM.</u> [folklore] Every closure system can be represented by an implicational base (at least one)

At least one ?



Def. Two implicational bases are equivalent if they represent the same closure system.

poly-time testable

Which one is the best ?

- (X,Σ) can enjoy minimality properties :

  (1) non-redundant : cannot remove any implication from Σ

  (2) minimum : Σ has the least possible number of implications — Our aim

  (3) optimum : $\sum_{A \to B \in \Sigma} |A| + |B|$ is minimal among all equiv. (X,Σ')

- (3) ⇒ (2) ⇒ (1) but (3) hard to optimize, while (1), (2) poly [Ausiello et al., 1986]

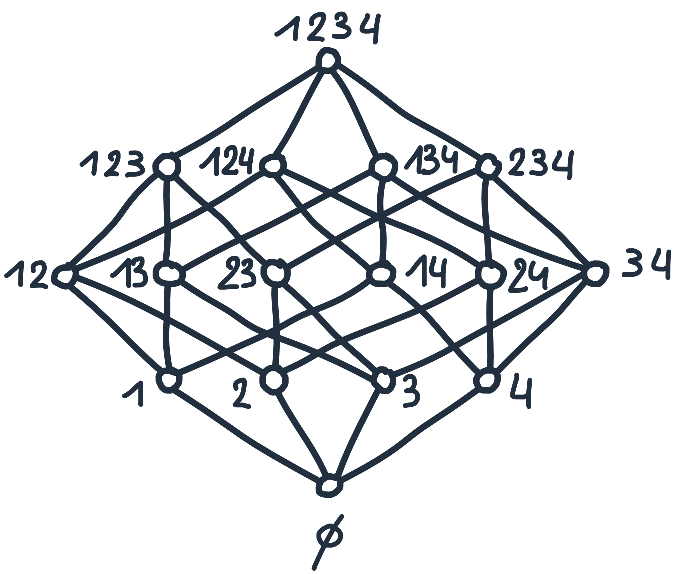
- (X,Σ) can have specific implications :

  minimum and easy to find !

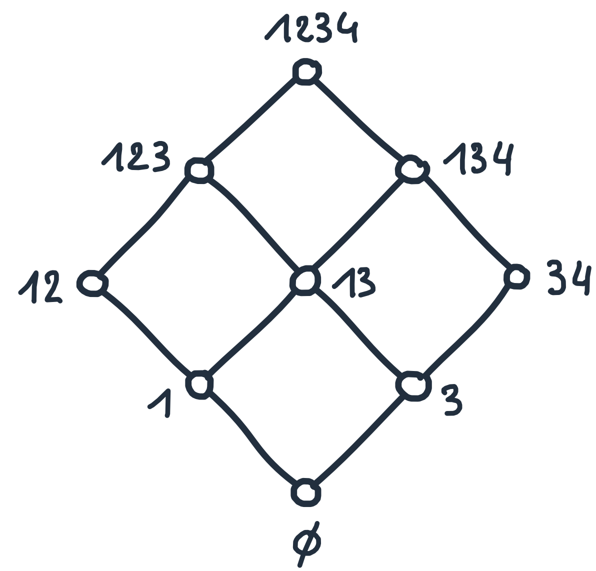
  - $P \to \phi(P)$ with P pseudo-closed ⇝ canonical base [Duquenne, Guigues, 1986]
  - $A \to b$ with A a minimal generator of b ⇝ canonical direct base [Bertet, Monjardet, 2010]

    $A \to b$ but $A' \not\to b$ for each $A' \subset A$
  - $A \to b$ with A a D-cover of b ⇝ D-base [Adaricheva et al., 2013]

Implicational bases: more examples



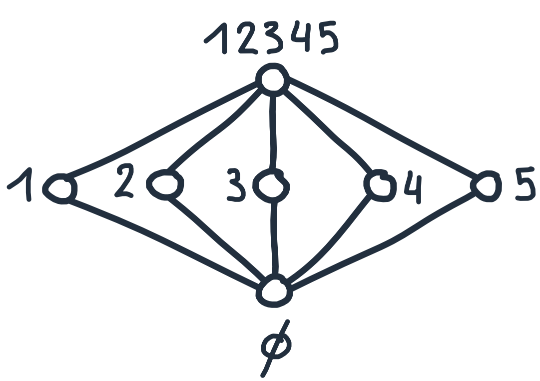
- $(X, \mathcal{C})$ Boolean cube, $|X| = n$
- $\Sigma = \emptyset$
- $\boxed{2^n = |\mathcal{C}| \gg |\Sigma| = n}$ exponential gap



K is fixed

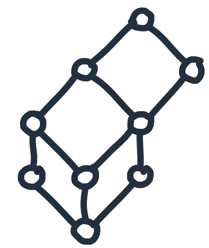
- $(X, \mathcal{C})$ is a $\boxed{\text{k-dim}}$ grid, $|X| = n$
- $\Sigma = \{x \to \phi(x) \mid \{x\} \notin \mathcal{C}, x \in X\}$
- $\boxed{n^k \approx |\mathcal{C}| > |\Sigma| \approx n}$

polynomial gap



- $(X, \mathcal{C})$ is a diamond, $|X| = n$
- $\Sigma = \{xy \to \phi(xy) \mid x, y \in X, x \neq y\}$
- $\boxed{n + 2 = |\mathcal{C}| < |\Sigma| \approx n^2}$ polynomial gap

Outline



closure systems

Same knowledge, different representations:
What is the complexity of translating between them?

Step 3. Answer the question

meet-irreducible elements

implicational bases

$2 \to 34$

$13 \to 4$

Formal statement

Prob. Implicational Base Identification (IBI)

- **In:** the meet-irreducible elements $\mathrm{Mi}(\mathcal{C})$ of a closure system $(X, \mathcal{C})$
- **Task:** find a minimum implicational base (X, Σ) for $(X, \mathcal{C})$

Prob. Computing Meet-Irreducible (CMI)

- **In:** an (minimum) implicational base (X, Σ) of a closure system $(X, \mathcal{C})$
- **Task:** find the meet-irreducible elements $\mathrm{Mi}(\mathcal{C})$ of $(X, \mathcal{C})$

- Hypothesis: $(X, \mathcal{C})$ is standard
  - $\emptyset \in \mathcal{C}$
  - $\boxed{\phi(x) \setminus \{x\} \in \mathcal{C} \text{ for all } x \in X}$ every x is useful

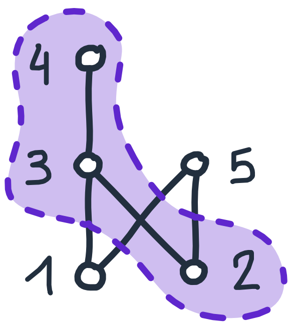
The problems are elsewhere

Finite sets	closure system	meet-irreducible	implications
KST	Knowledge space	atoms	queries / entailment
FCA	concept lattice	(reduced) context	attribute implications
Horn logic	models	Characteristic models	Pure Horn CNF
Databases	Closure system	Armstrong relation	Functional Dependencies

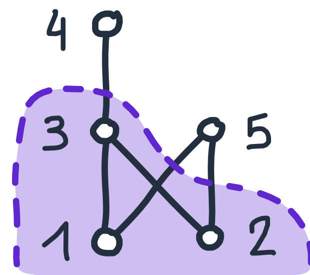
Depending on the field, one of IBI or CMI is more natural

- queries : "If the students fail the items in A, they fail the items in B"
- attribute implications : "The objects having attributes A also have attributes B"
- Horn CNF : Horn clause $(\bar{1} \vee \bar{3} \vee 4) \leftrightarrow$ implication $13 \rightarrow 4$
- Functional Dependencies : "Two tuples equal on A are equal on B"

Other sources of closure systems

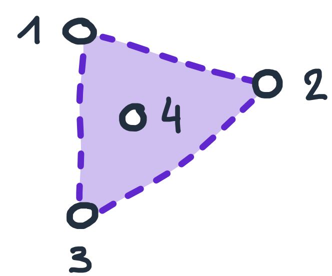


- Poset
- Convex sets
- 24 → 3

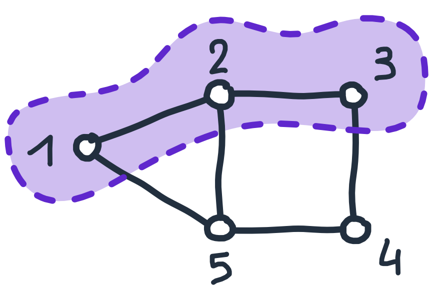


- Poset
- Ideals (down-set)
- 3 → 12

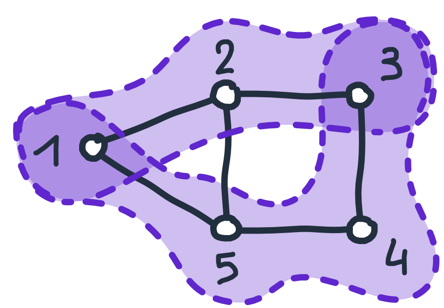
Closure systems arise from several objects
→ particular cases of CMI, IBI



- Points in $\mathbb{R}^n$
- Convex hull
- 123 → 4



- Graph
- shortest path
- 13 → 2



- Graph
- induced path
- 13 → 2, 13 → 45

Comparing the representations

Question	(X, Σ)	$Mi(\mathcal{C})$	$(X, \mathcal{C})$
is x in a minimal key ?	NP-c	poly	poly
is P pseudo-closed ?	poly	coNP-c	poly
is $\mathcal{C}$ a convex geometry ?	NP-c	poly	poly
Relative size	Brand new ! [Adaricheva, Bichoupan, 2023]		
size of ... w.r.t. Σ	–	$\exp(\lvert\Sigma\rvert)$	$\exp(\lvert\Sigma\rvert)$
size of ... w.r.t. $Mi(\mathcal{C})$	$\exp(\lvert Mi(\mathcal{C})\rvert)$	–	$\exp(\lvert Mi(\mathcal{C})\rvert)$
size of ... w.r.t. $\mathcal{C}$	$\leqslant \lvert\mathcal{C}\rvert \times \lvert X\rvert$	$\leqslant \lvert\mathcal{C}\rvert$	–

Each representation, Σ or Mi, can be much smaller than the other

The complexity of a problem depends on the representation

Packing up motivations

- Why studying IBI and CMI ?
  - The problems arise from different fields
  - They are impacted by the type of closure system at hand
  - Each representation has its own benefits

$\boxed{\text{So now ... what about their complexity ?}}$

Enumeration: idea

NOT count

- IBI and CMI are enumeration problems: we try to list objects
- But the output may have exponential size w.r.t. the input ...

Prob. Powerset

In: a set X

Task: list all the subsets of X

- Powerset is easy to solve
- But any algorithm will take at least $O(2^{|X|})$ time ...

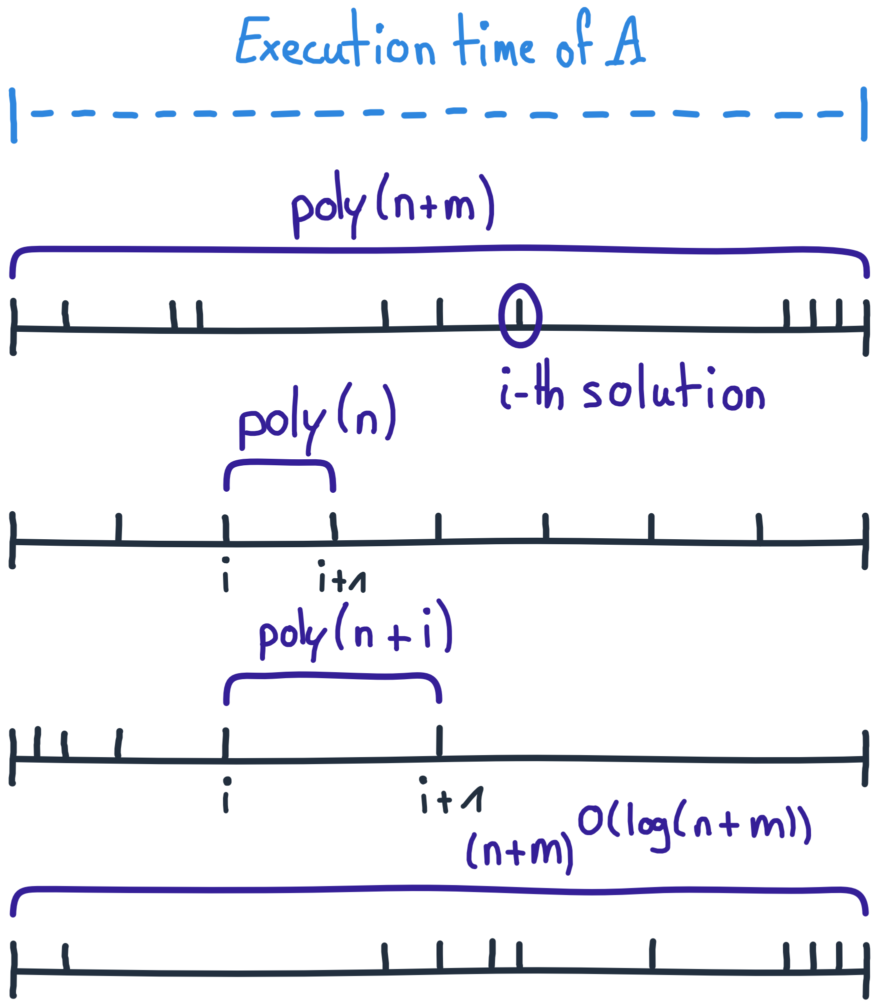
IDEA: take output size into account → output-sensitive complexity (see e.g. [Johnson et al., 1988])

- Powerset can be solved in poly-time in its input and output

$|X|$ $\quad 2^{|X|}$

Enumeration: output-sensitive complexity

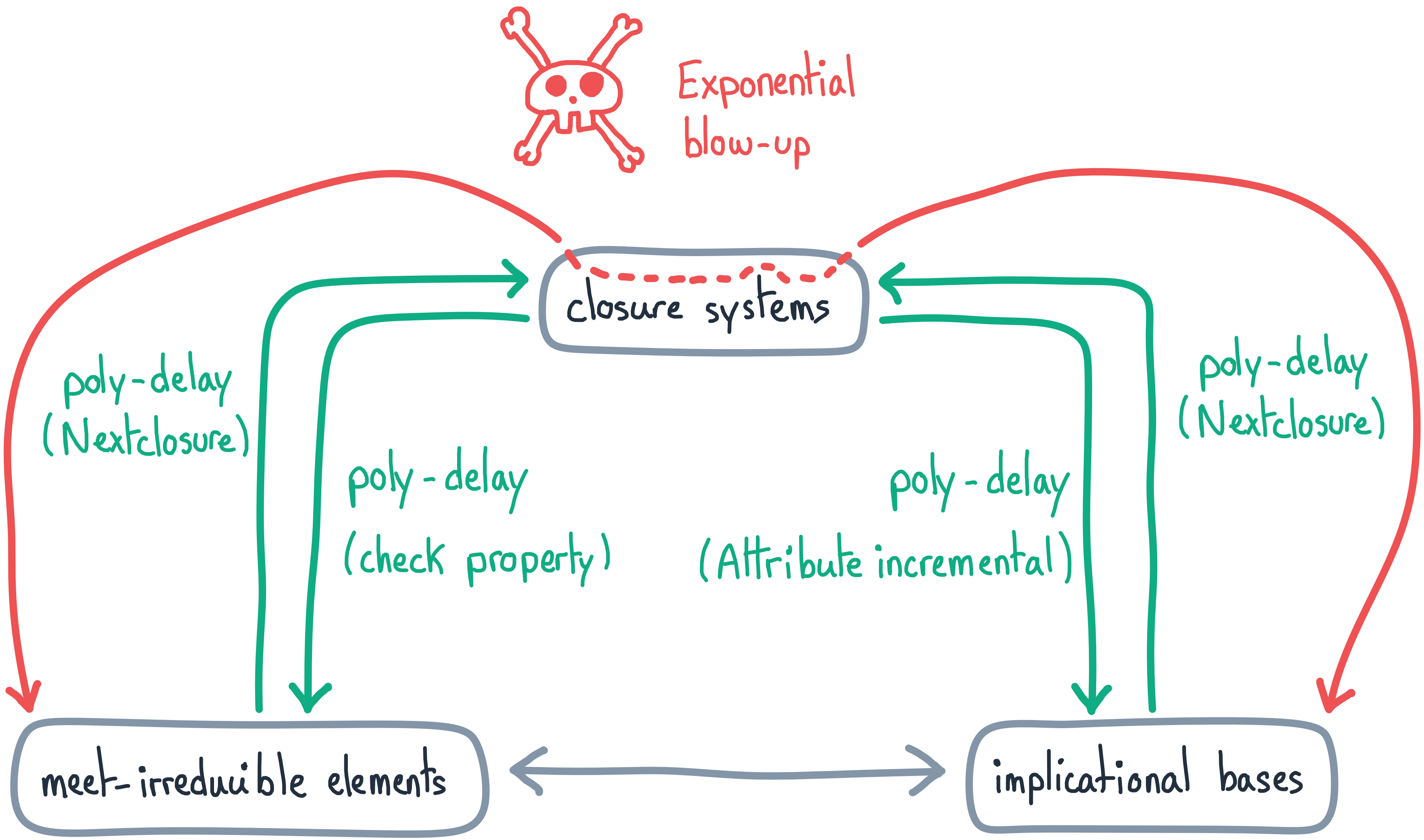
- Enumeration task: given an input x, list a set of $\boxed{\text{solutions}}$ $R(x)$ — of size poly(x)



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

- output-polynomial time
- polynomial delay
- incremental polynomial time
- output quasi-polynomial time

First idea



Cannot compute the closure system to solve IBI (or CMI)

The truth

The complexity of CMI and IBI is unknown ...

- Harder than enumerating the maximal independent sets of a hypergraph [Khardon, 1995]

  let's see this !

- Finding the maximal meet-irreducible elements is hard [Kavvadias et al., 2000]
- Finding the maximal pseudo-closed sets is hard [Babin, Kuznetsov, 2013]

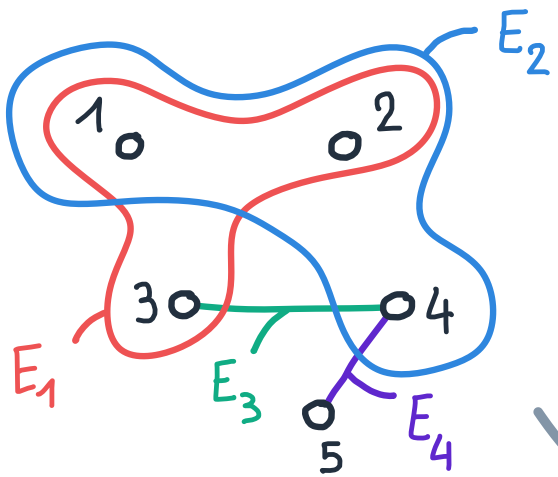
−

- General (exponential) algorithms [Mannila, Räihä, 1992], [Wild, 1995]
- Tractable cases : $SD_\wedge$ lattices, types of convex geometries, modular lattices, ...
  [Beaudou et al., 2017], [Nourine, V., 2023+], [Wild, 2000]

+

- Surveys [Bertet et al. 2018], [Wild, 2017]

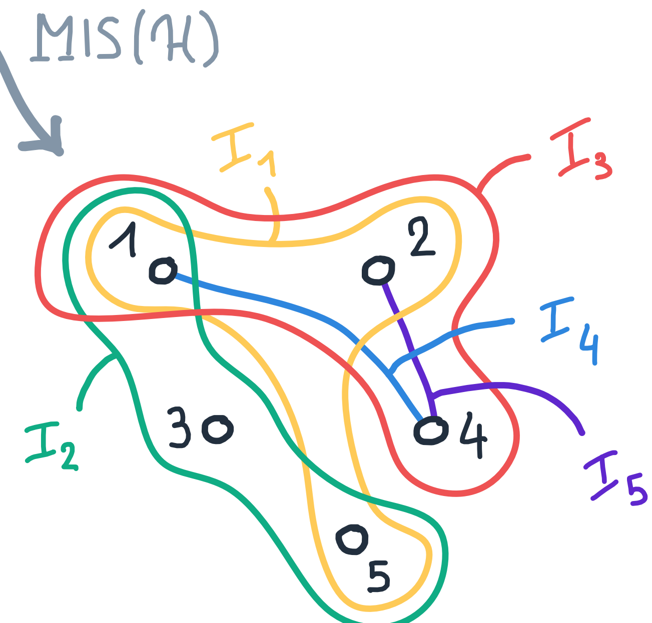
Hypergraphs



Def. A hypergraph is a pair $\mathcal{H}=(X,\mathcal{E})$ where X is a set and $\mathcal{E}$ a collection of subsets of X

- $\mathcal{H} = (X = \{1,\dots,5\}, \{E_1, E_2, E_3, E_4\})$:

  $E_1 = 123$, $E_2 = 124$, $E_3 = 34$, $E_4 = 45$



Def. Let $\mathcal{H}=(X,\mathcal{E})$ be a hypergraph. A set $I \subseteq X$ is an independent set of $\mathcal{H}$ if $E \not\subseteq I$ for all $E \in \mathcal{E}$

- Maximal ($\subseteq$) independent sets of $\mathcal{H}$

  $\mathrm{MIS}(\mathcal{H}) = \{I_1, I_2, I_3, I_4, I_5\}$

  $I_1 = 125$, $I_2 = 135$, $I_3 = 124$, $I_4 = 14$, $I_5 = 24$

PROB. Enum Max. Ind. Sets (Enum-MIS)

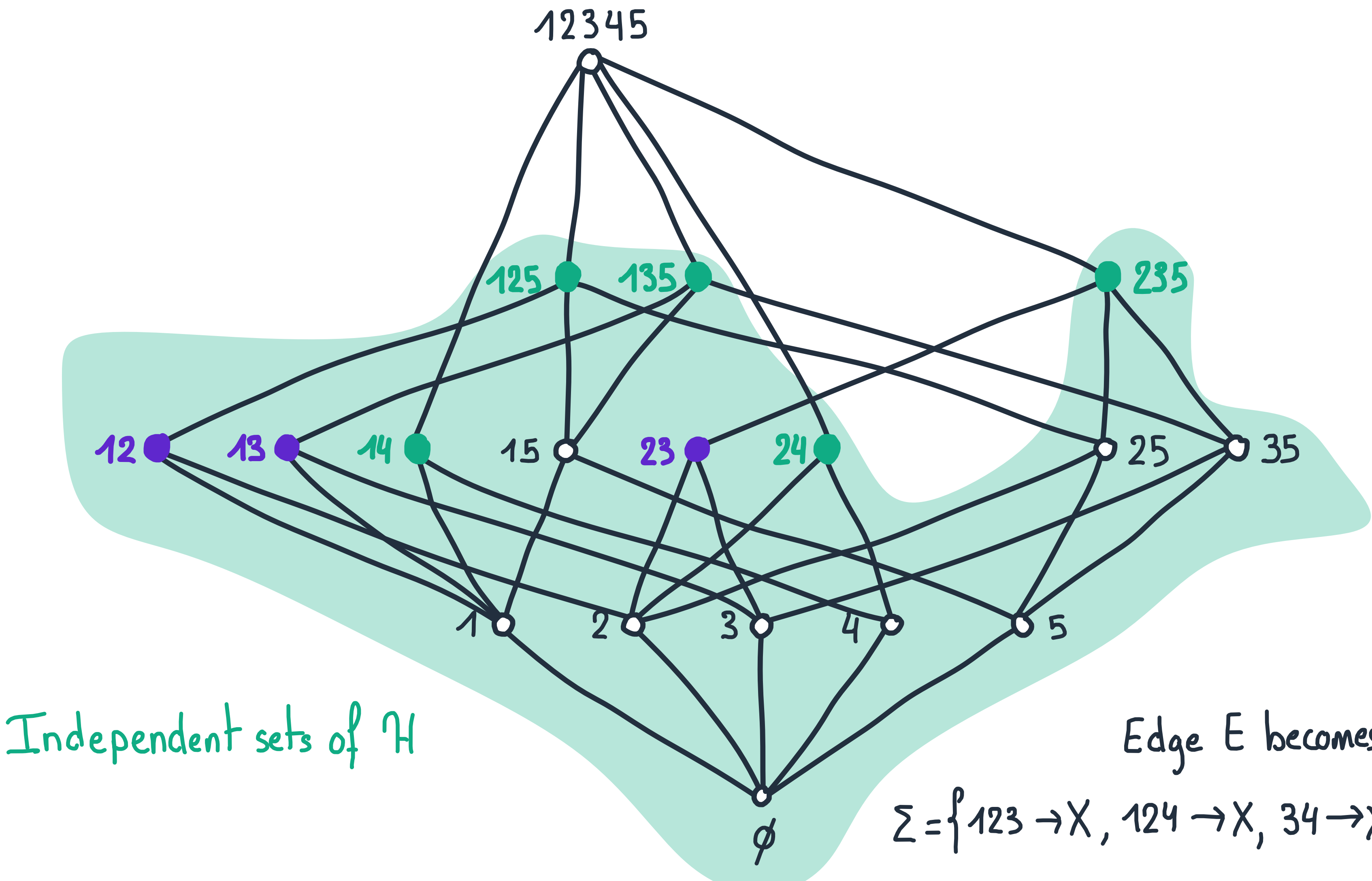
In: a simple hypergraph $\mathcal{H} = (X, \mathcal{E})$ $E_i \nsubseteq E_j \; \forall E_i, E_j \in \mathcal{E}$

Task: enumerate the maximal ($\subseteq$) independent sets of $\mathcal{H}$, $MIS(\mathcal{H})$

- Open problem... [Eiter et al., 2008]
- quasi-poly algorithm [Fredman, Khachiyan, 1996]

CMI is harder than Enum-MIS

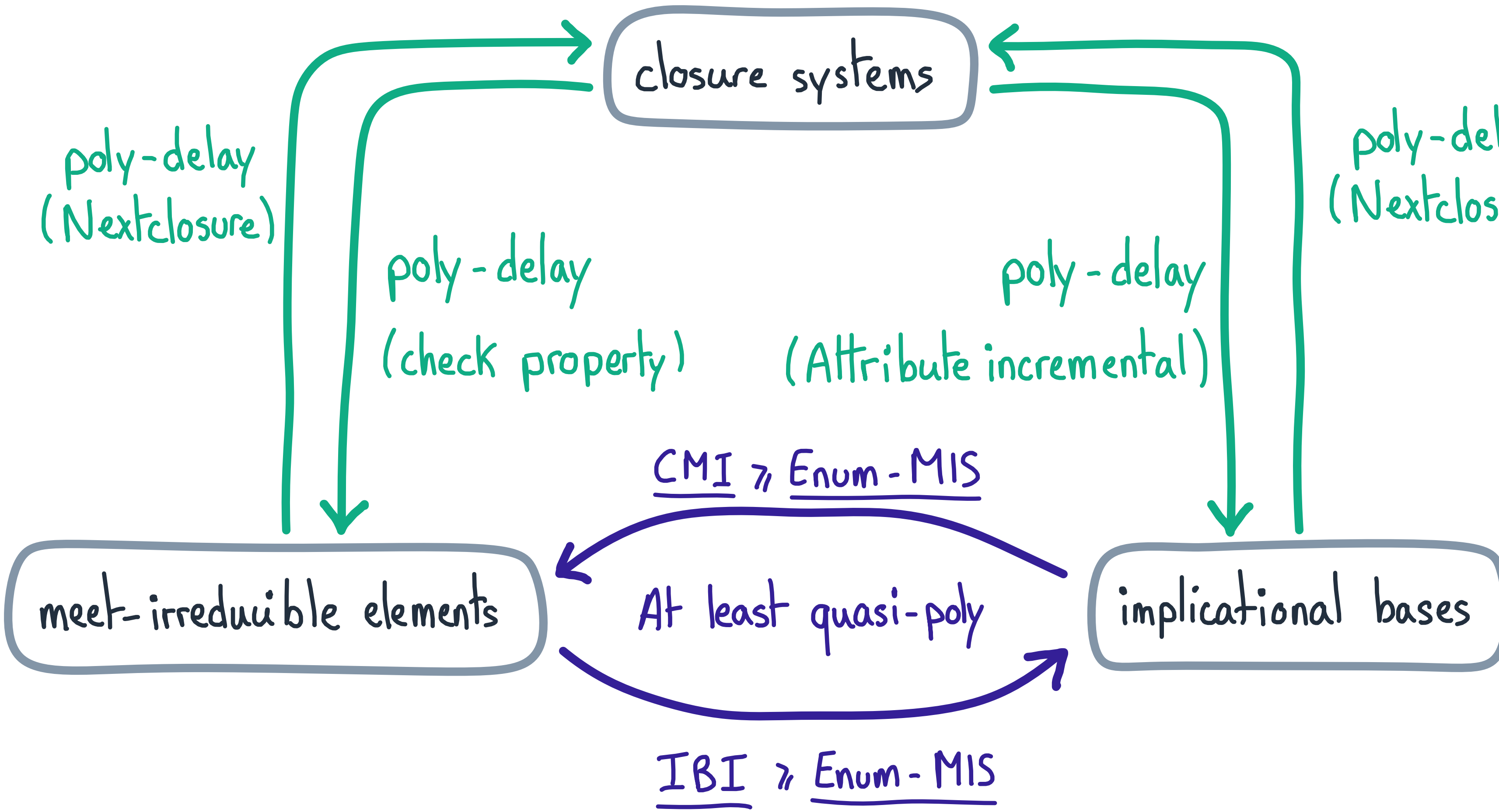
$\mathrm{Mi}(\mathcal{E}) = \mathrm{MIS}(\mathcal{H}) + \text{Some of the } I \setminus \{x\},\ I \in \mathrm{MIS}(\mathcal{H})$



Edge E becomes $E \to X$

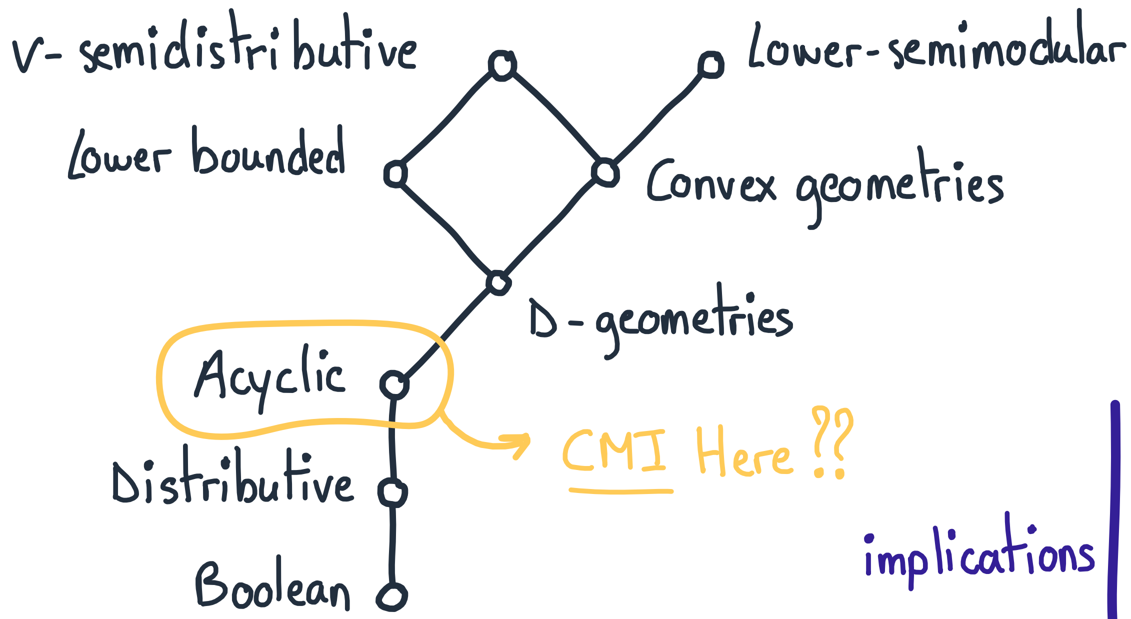
$\Sigma = \{123 \to X,\ 124 \to X,\ 34 \to X,\ 45 \to X\}$

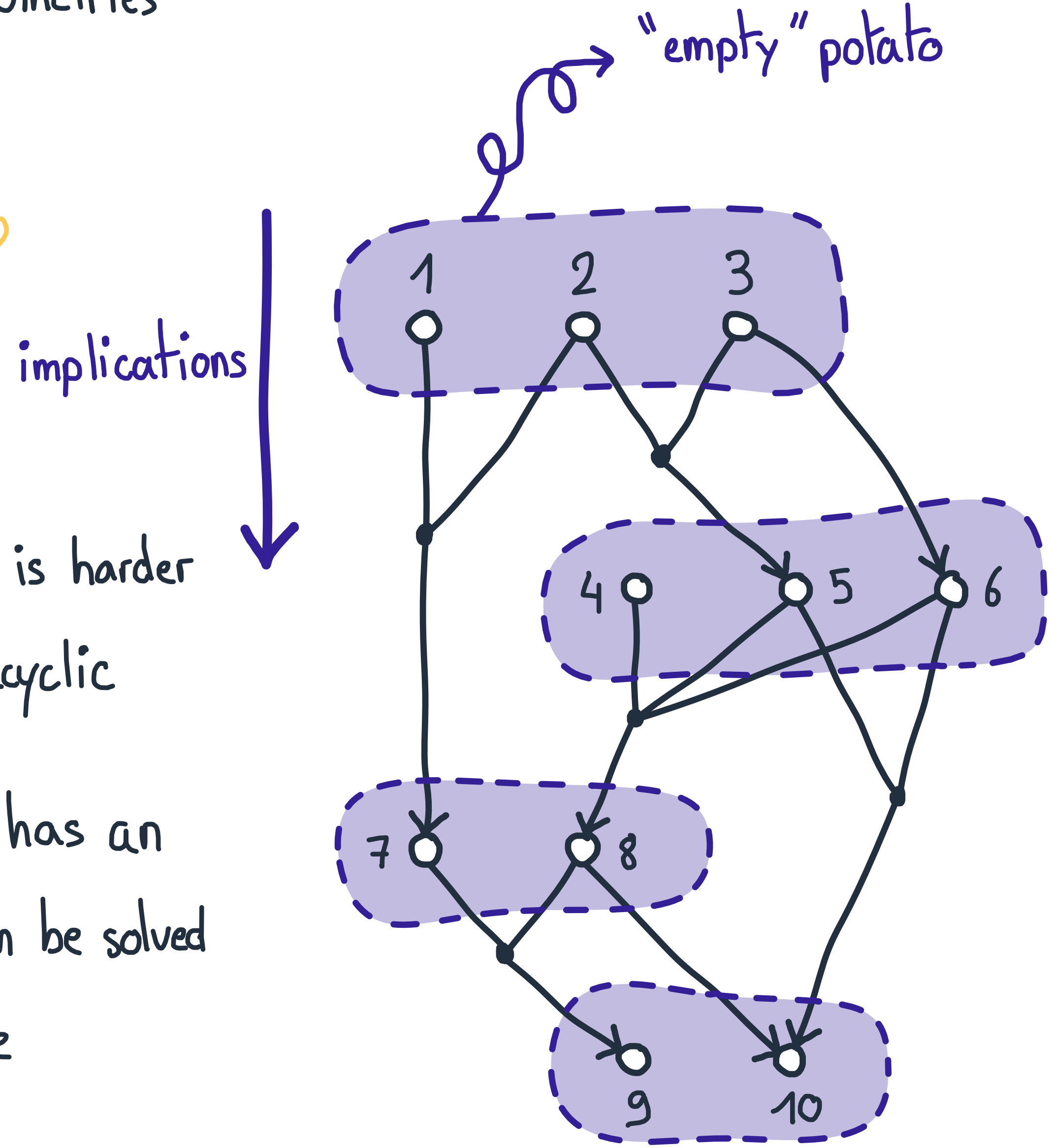
The big picture



Thm. [Khardon, 1995] CMI and IBI are harder than Enum-MIS

A glimpse of our results





THM. [Defrain, Nourine, V., 2021] CMI is harder than Enum-MIS even if (X, Σ) is acyclic

THM. [Nourine, V., 2023+] If (X, Σ) has an appropriate decomposition, CMI can be solved in output quasi-polynomial time

Conclusion

- Closure systems are ubiquitous but huge and complex $\longrightarrow$ use representations !
  - Implications : " If we have A, we have B "
  - Meet-irreducible elements : the core of the system

- Translating between the representations ($\underline{\text{CMI}}$, $\underline{\text{IBI}}$) is fascinating but tough

- Our progress on acyclic implications
  - the problem ($\underline{\text{CMI}}$) is already quite hard ($\geq$ $\underline{\text{Enum-MIS}}$)
  - But we can manage some cases with nice decompositions

- What's next ? Acyclic and beyond ? Other classes ? Hardness of the problem ?

References

- Doignon, Falmagne [Doignon, Falmagne, 1985]
  Spaces for the assessment of Knowledge
  International Journal of man-machine studies, 1985
- Ausiello, D'Atri, Saccà [Ausiello et al., 1986]
  Minimal representations of directed hypergraphs
  SIAM Journal on Computing, 1986
- Guigues, Duquenne [Duquenne, Guigues, 1986]
  Familles minimales d'implications informatives résultant d'un tableau de données binaires
  Mathématiques et Sciences humaines, 1986
- Bertet, Monjardet [Bertet, Monjardet, 2010]
  The multiple facets of the canonical direct base
  Theoretical Computer Science, 2010
- Adaricheva, Bichoupan [Adaricheva, Bichoupan, 2023]
  Complexity results for implication bases of convex geometries
  arXiv, 2023

References

- Johnson, Yannakakis, Papadimitriou [Johnson et al., 1988]
  On generating all maximal independent sets
  Information Processing Letters, 1988
- Khardon [Khardon, 1995]
  Translating between Horn representations and their Characteristic Models
  Journal of Artificial Intelligence Research, 1995
- Kavvadias, Sideri, Stavropoulos [Kavvadias et al., 2000]
  Generating all maximal models of a Horn expression
  Information Processing Letters, 2000
- Babin, Kuznetsov [Babin, Kuznetsov, 2013]
  Computing premises of a minimal cover of functional dependencies is intractable
  Discrete Applied Mathematics, 2013
- Mannila, Räihä [Mannila, Räihä, 1992]
  The Design of Relational Databases
  Addison-Wiley, 1992

References

- Wild [Wild, 1995]
  Computations with finite closure systems and Implications
  Springer LNCS 959, 1995

- Beaudou, Mary, Nourine [Beaudou et al., 2017]
  Algorithms for k-meet-semidistributive lattices
  Theoretical Computer Science, 2017

- Nourine, Vilmin [Nourine, V., 2023+]
  The enumeration of meet-irreducible elements based on hierarchical decompositions of implicational bases
  arXiv, 2023

- Wild [Wild, 2000]
  Optimal implicational bases for finite modular lattices
  Quaestionas Mathematicae, 2000

- Bertet, Demko, Viaud, Guérin [Bertet et al. 2018]
  Lattices, closure systems and implication bases: A survey of structural aspects and algorithms
  Theoretical Computer Science, 2018

References

- Adaricheva, Nation, Rand [Adaricheva et al., 2013]
  Ordered direct implicational basis of a finite closure system
  Discrete Applied Mathematics, 2013
- Wild [Wild, 2017]
  The joy of implications, aka pure Horn formulas: mainly a survey
  Theoretical Computer Science, 2017
- Eiter, Makino, Gottlob [Eiter et al., 2008]
  Computational aspects of monotone dualization: a brief survey
  Discrete Applied Mathematics, 2008
- Fredman, Khachiyan [Fredman, Khachiyan, 1996]
  On the complexity of dualization of monotone disjunctive normal forms
  Journal of Algorithms, 1996
- Defrain, Nourine, Vilmin [Defrain, Nourine, V., 2021]
  Translating between the representations of a ranked convex geometry
  Discrete Mathematics, 2021.