Enumerating maximal consistent closed sets in closure systems

ICFCA 2021, International Conference on Formal Concept Analysis

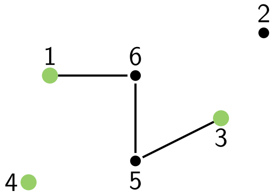
Lhouari Nourine and *Simon Vilmin*.

LIMOS, UCA

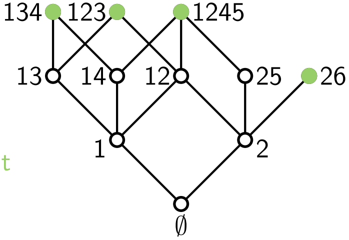
June 30th, 2021

Implications and inconsistency

- A set X of attributes with implications Σ, and closure system $\mathcal{F}$
- A *consistency graph* G (over X)
- Problem MCCEnum : enumerate *maximal consistent closed sets* of $\mathcal{F}$ w.r.t G .



134 max consistent



Formally

Problem - Maximal consistent closed sets enumeration (MCCEnum)

- **Input:** A set of implications Σ over X, a consistency graph $G = (X, E)$.
- **Output:** maximal consistent closed sets of $\mathcal{F}$ w.r.t G, denoted $\text{maxCC}(\Sigma, G)$.

Origins:

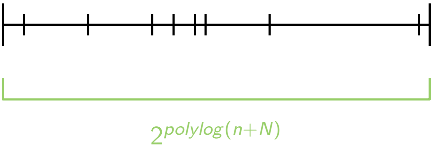
- Representation for median-semilattices [Barthélemy, Constantin, 1993], [Nielsen et al., 1981]
- Extended to modular-semilattices with applications to combinatorial optimization [Hirai, Nakashima, 2020], [Hirai, Nakashima, 2018].

- MCCEnum *output-polynomial* for modular/median-semilattices [Hirai, Nakashima, 2018], [Kavvadias et al, 1995]
- Restricted case of dualization in lattices given by implicational bases, an *NP-complete problem* [Kavvadias et al, 1995], [Babin, Kuznetsov, 2017].

Quick recap on enumeration

- In enumeration, the size N of the output may be *exponential* in the input size n.
- Let $\mathcal{A}$ be an enumeration algorithm
  - Execution time bounded by poly$(n+N)$: $\mathcal{A}$ runs in *output-polynomial* time
  - Delay poly(n) between two outputs: $\mathcal{A}$ has *polynomial delay*
  - Delay poly$(n+i)$ between i-th and $i+1$-th outputs: $\mathcal{A}$ runs in *incremental-polynomial* time
  - Execution time bounded by $2^{\text{polylog}(n+N)}$: $\mathcal{A}$ runs in *output-quasipolynomial* time.

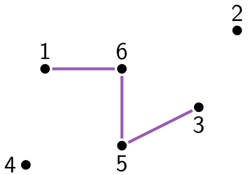
Execution time of $\mathcal{A}$

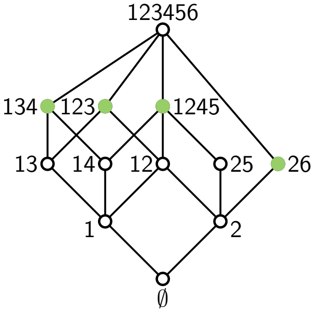


Connexion with co-atoms

- Maximal consistent closed sets are now *co-atoms*.

$$\Sigma = \left[\begin{array}{ll} 6 \to 2, & 16 \to 5, \\ 5 \to 2, & 15 \to 4, \\ 3 \to 1, & 24 \to 5, \\ 4 \to 1, & 16 \to X, \\ 56 \to X, & 35 \to X. \end{array} \right]$$





Connexion with co-atoms (bis)

- When adding edges of G as *keys*
- MCCEnum becomes a *restricted* instance of :

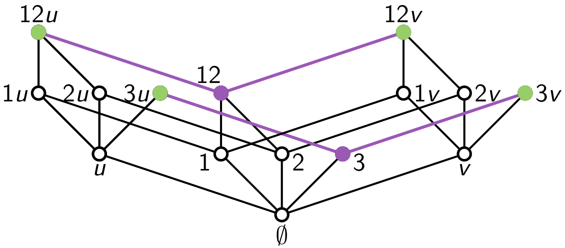
Problem - Co-atoms enumerations (CE)

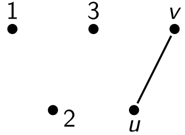
- **Input:** A set of implications Σ over X.
- **Output:** Co-atoms of $\mathcal{F}$.

- But, CE is *untractable* in output-poly time (unless P $=$ NP) ! [Kavvadias et al, 1995]
- So, what about our problem ?

Proof by picture

- Start from Σ over X (with induced $\mathcal{F}$)
- Create fresh elements u, v, add $X \rightarrow uv$ to Σ
- The graph G (over $X \cup \{u, v\}$) has a unique edge: uv
- *maximal consistent closed sets* are duplications of $\mathcal{F}$'s *co-atoms*.





Results

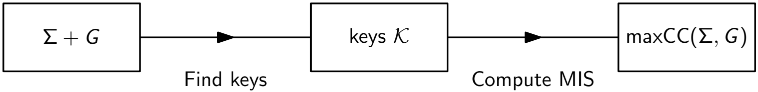
Theorem [Nourine, V., 2021+]: The problem MCCENUM cannot be solved in output-polynomial time unless P = NP.

- Look carefully at the reduction of [Kavvadias et al, 1995]
- Slightly change the implication $X \rightarrow uv$
- Use the D-relation of (see e.g. [Freese et al, 1995]) for lower bounded closure systems.

Corollary [Nourine, V., 2021+]: The problem MCCENUM cannot be solved in output-polynomial time unless P = NP, even when restricted to lower bounded closure systems.

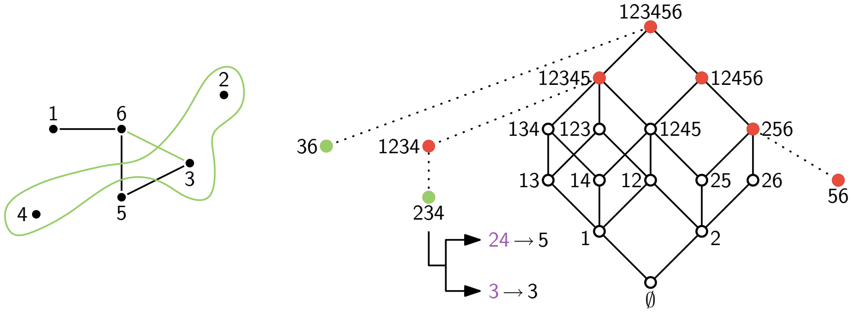
An approach to tractable cases

- Co-atoms of are maximal independent sets of keys:
  - Add to Σ the implications $uv \rightarrow X$, for edges uv of G
  - Compute the keys $\mathcal{K}$
  - Find the maximal independent sets $\mathrm{MIS}(\mathcal{K}) = \mathrm{maxCC}(\Sigma, G)$.



How do keys behave ?

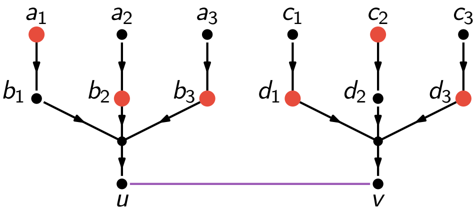
- In our case, a subset F must contain a key if:
  - it *contains* an edge of G,
  - its *closure contains* an edge of G
- Thus, a *key* K has the form $K = A_u \cup A_v$ with A_u, A_v minimal generators of u, v
  - A_u *minimal generator* of u if it is a minimal subset of X such that $A_u \rightarrow u$.



Exponential Example

- $X = \{a_1, \ldots, a_n, b_1, \ldots, b_n, c_1, \ldots, c_n, d_1, \ldots, d_n, u, v\}$
- Σ with implications:

- G has a unique *edge* uv
- $\text{maxCC}(\Sigma, G)$ has $2n$ *solutions* : X minus a triple $\{u, a_i, b_i\}$ or $\{v, c_i, d_i\}$
- at least 2^{2n} *keys* of *size* $2n$: all binary words on $\{a_i, b_i\}^n \times \{c_i, d_i\}^n$.

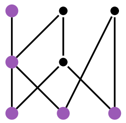


Carathéodory number

- Carathéodory number $c(\mathcal{F})$: max. size of a minimal generator
- If bounded by a constant k:
  - Keys have size $\leq 2 \times k$
  - and $\mathcal{K}$ has size poly($|X|$) !
- MIS($\mathcal{K}$) = maxCC(Σ, G) computable in incremental-poly time [Eiter et al., 1996]

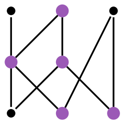
Theorem [Nourine, V., 2021+]: If the Carathéodory number is bounded by a constant, MCCEnum can be solved in incremental-polynomial time.

Some closures with bounded c($\mathcal{F}$)



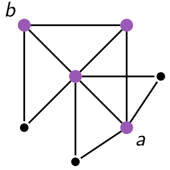
ideals of a poset

$c(\mathcal{F}) = 1$



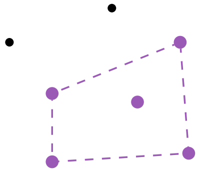
convex subsets of a poset

$c(\mathcal{F}) = 2$



monophonic convexity of a chordal graph

$c(\mathcal{F}) = 2$

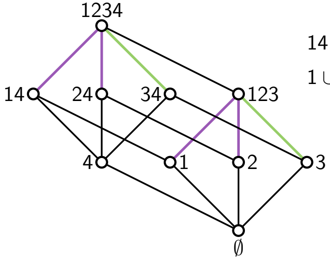


convex hull in $\mathbb{R}^k$

$c(\mathcal{F}) = k + 1$

The (atomistic) modular case

- MCCENUM easy to solve in distributive closure systems
- What about modularity ? Focus on the atomistic case:
  - *independence* of minimal generators [Grätzer, 2011]: subsets of A_u generates a boolean sublattice of $\mathcal{F}$
  - *biatomicity* [Bennett, 1987]: if $F \in \mathcal{F}$ and $F \cup x \rightarrow y$, there exists $z \in F$ such that $zx \rightarrow y$
  - $\implies$ $c(\mathcal{F}) \leq \log_2(|X|)$
- So, keys have size *at most* $2 \times \log_2(|X|)$!



$14 \cup 2 \rightarrow 3$

$1 \cup 2 \rightarrow 3$

Theorem [Nourine, V., 2021+]: The problem MCCENUM can be solved in

MCCEnum : The big picture

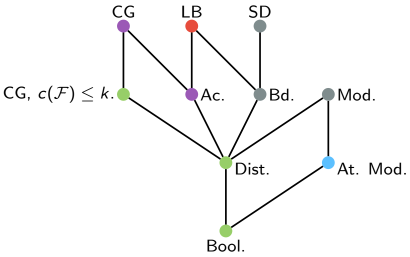
output-poly

quasi-poly

$\mathcal{K}$ exponential

untractable

unknown



Bool. = Boolean

Dist. = Distributive

At. = Atomistic

Mod. = Modular

Ac. = Acyclic

CG = Convex Geometry

Bd. = Bounded

LB = Lower Bounded

SD = Semidistributive

Conclusion

- Problem: given Σ and $G = (X, E)$, find maximal consistent (i.e independent) closed sets of $\mathcal{F}$ w.r.t to G
- Results:
  - ✗ Not solvable in output-polynomial time unless P = NP,
  - ✓ Incremental-polynomial if the Carathéodory number is bounded,
  - ✓ Output-quasipolynomial in atomistic modular closure systems.
- Further research:
  - ▷ Tractability if the context is given as an input ?
  - ▷ Output-polynomial classes of closure systems generalizing distributivity ?

References

- J-P. Barthélemy, J. Constantin
  Median graphs, parallelism and posets.
  Discrete Mathematics, 111 :49-63, 1993.
- M. Nielsen, G. Plotkin, G. Winskel
  Petri nets, event structures and domains, part I.
  Theoretical Computer Science, 13 :85-108, 1981.
- H. Hirai, S. Nakashima
  A compact representation for modular semilattices and its applications.
  Order, 37 :479-507, 2020.
- H. Hiroshi and T. Oki
  A compact representation for minimizers of k-submodular functions.
  Journal of Combinatorial Optimization, 36 :709-741, 2018.
- T. Eiter, G. Gottlob, K. Makino.
  New results on monotone dualization and generating hypergraph transversals.
  SIAM Journal on Computing, 32 :514-537, 2003.
- M. Fredman, L. Khachiyan.
  On the complexity of dualization of monotone disjunctive normal forms.
  Journal of Algorithms, 21 :618-628, 1996.

References

- D. Kavvadias, M. Sideri, E. Stavropoulos
  Generating all maximal models of a Boolean expression.
  Information Processing Letters, 74 :157-162, 2000.

- M. Babin, S. Kuznetsov
  Dualization in lattices given by ordered sets of irreducibles.
  Theoretical Computer Science, 658 :316-326, 2017.

- L. Nourine, S. Vilmin.
  Enumerating maximal consistent closed sets in closure systems
  arXiv preprint arXiv:2102.04245, 2021.

- M.K. Bennett
  Biatomic lattices.
  Algebra Universalis, 24 :60-73, 1987.

- G. Grätzer
  Lattice theory: foundation.
  Springer Science & Business Media, 2011.

- R. Freese, J. Ježek, J.B. Nation
  Free lattices.
  American Mathematical Soc., 1995.