Improving Term Extraction with Acyclic Constraints

Mike He, Haichen Dong, Sharad Malik and Aarti Gupta

EGRAPHS 2023

PRINCETON UNIVERSITY

Overview

+ Term Extraction
+ Extraction with ILP and its challenge
+ Our contribution: Acyclic constraints

Term Extraction

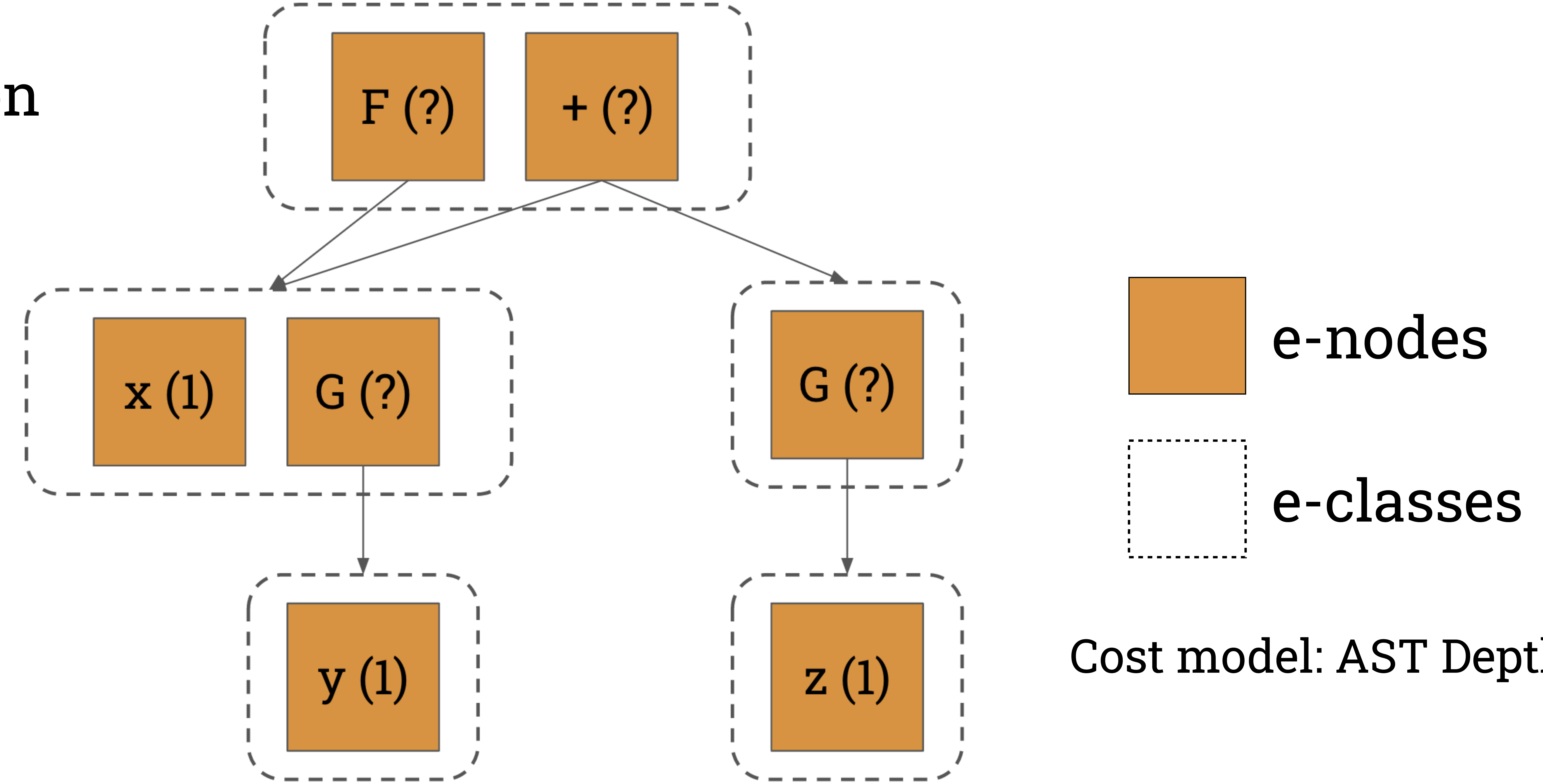
<u>Inputs:</u> e-graph; root e-class(es); cost model

<u>Output:</u> term(s) w/ the minimum cost

Term Extraction

Greedy Extraction

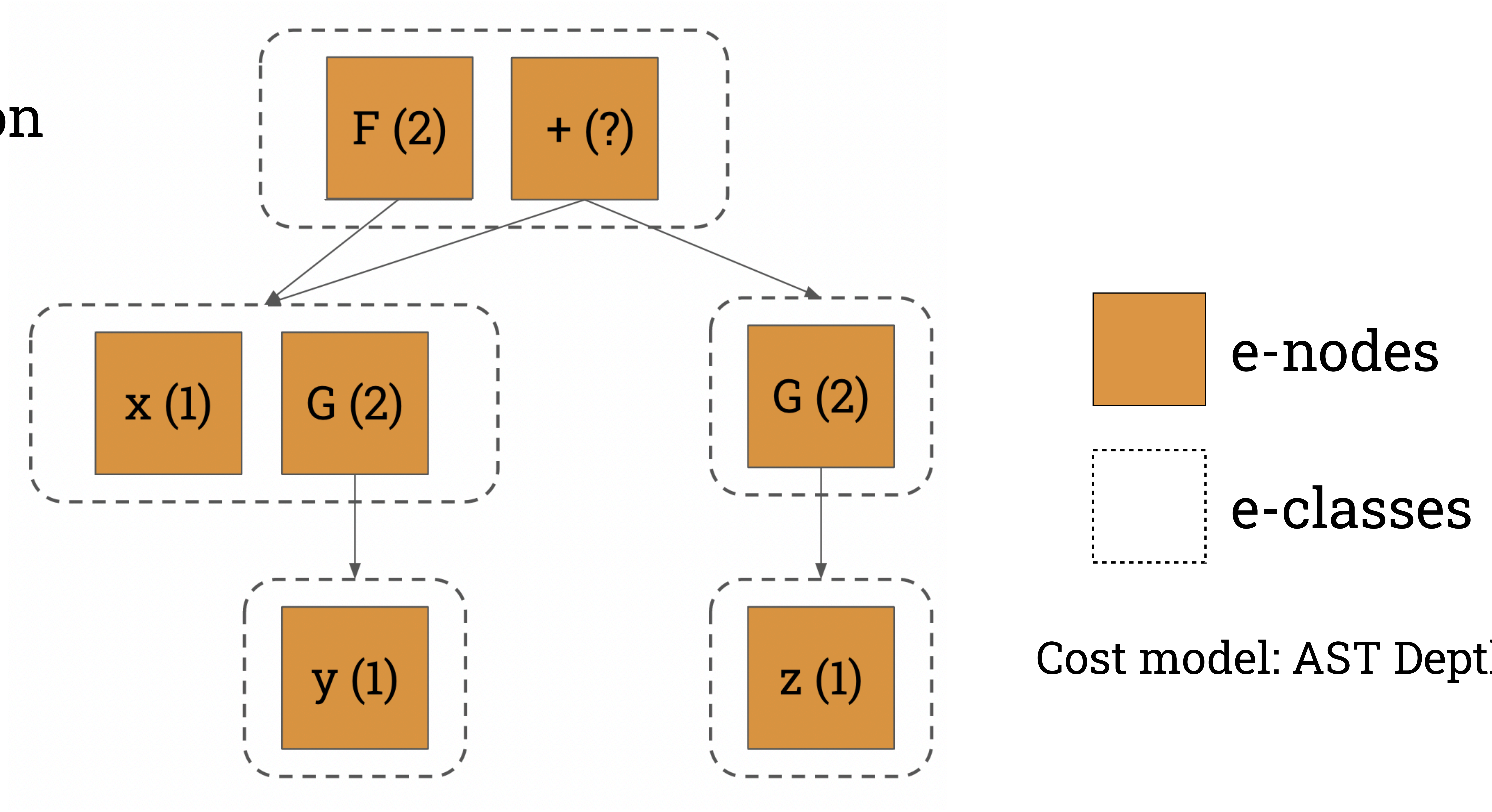
P1: Build costs from bottom-up



Cost model: AST Depth

Term Extraction

Greedy Extraction



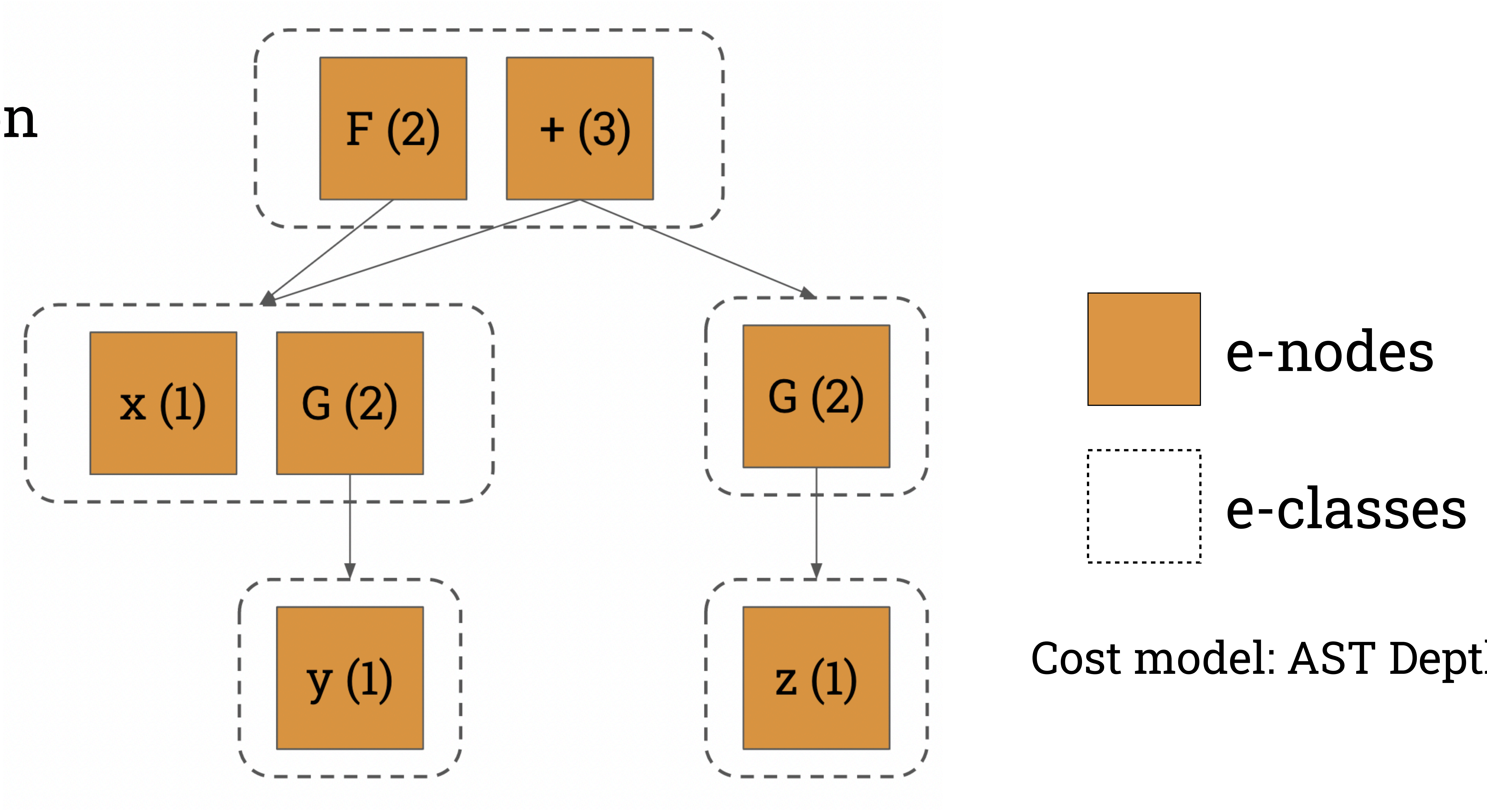
P1: Build costs from bottom-up

Cost model: AST Depth

Term Extraction

Greedy Extraction

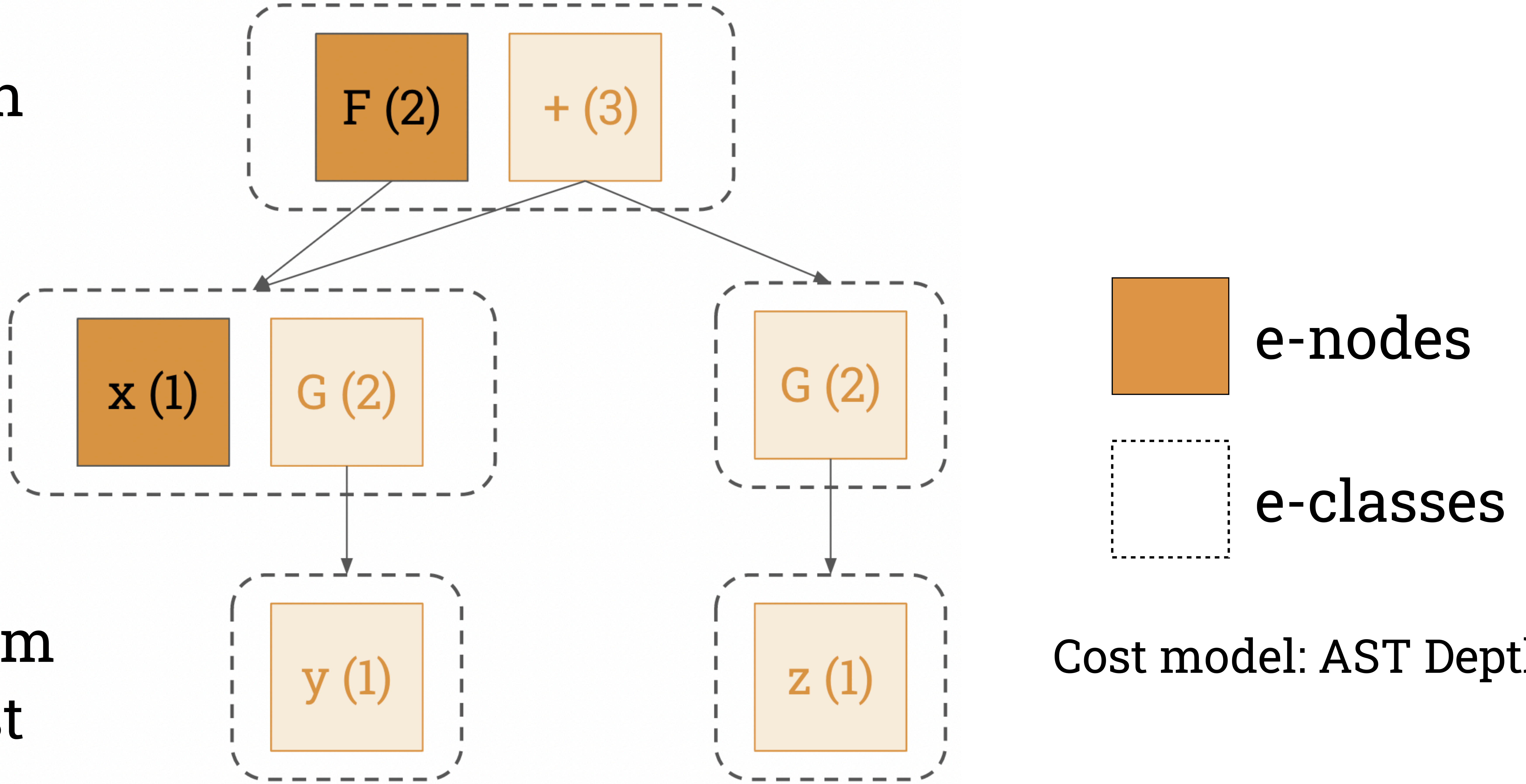
P1: Build costs from bottom-up



Cost model: AST Depth

Term Extraction

Greedy Extraction



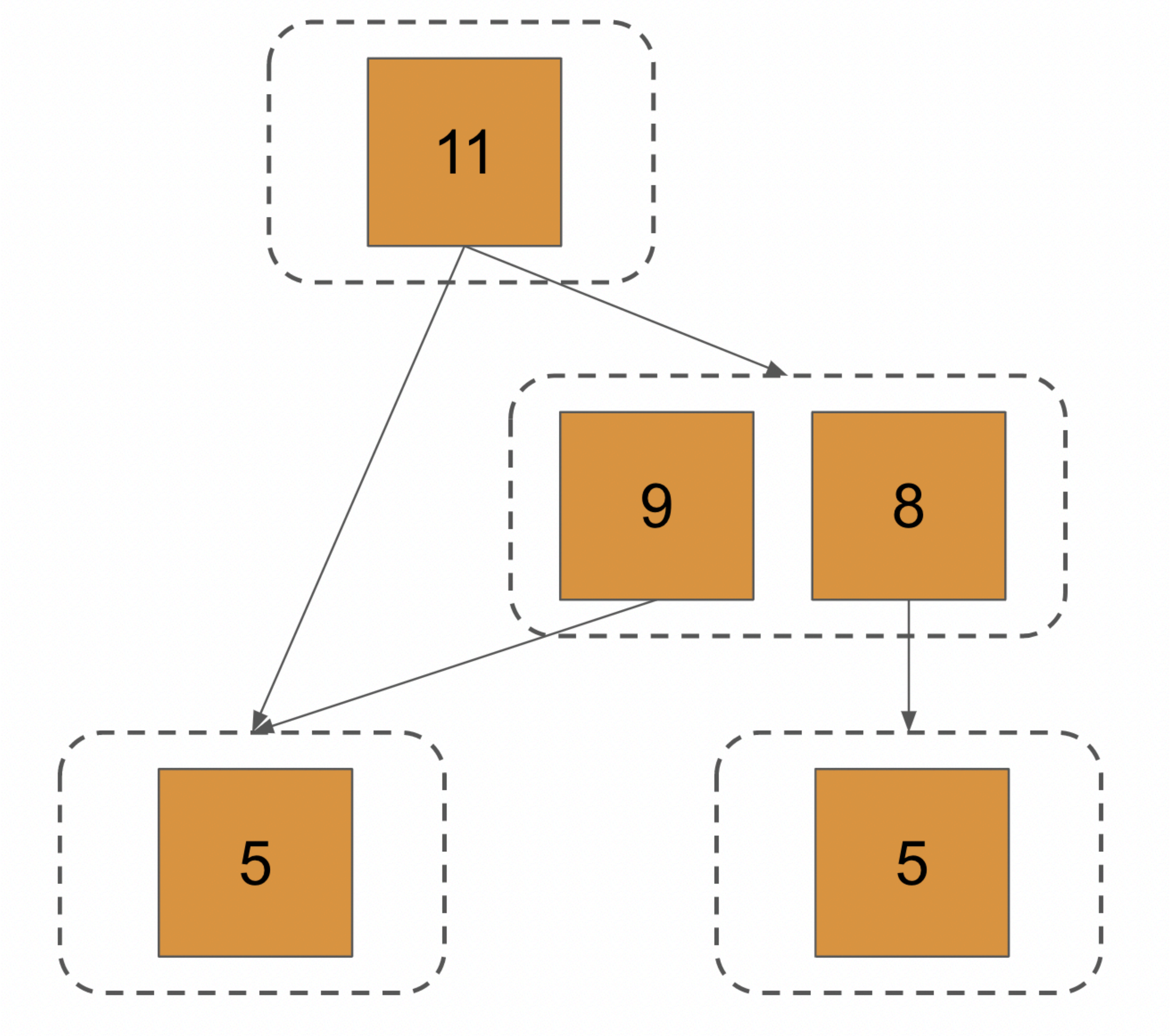
P2: choose the term with the least cost

Cost model: AST Depth

Term Extraction

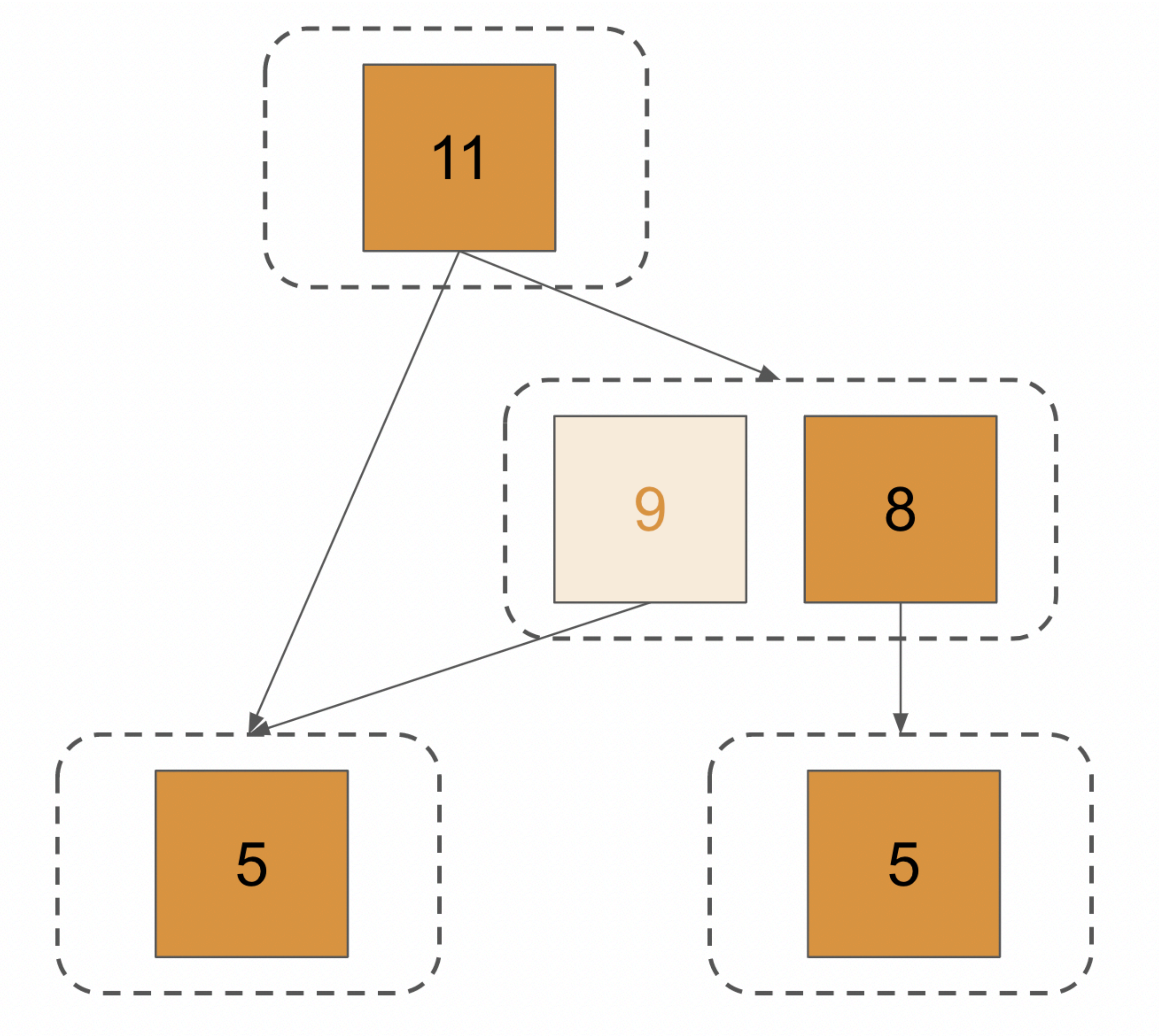
Greedy works in most cases, but may fail by yielding sub-optimal solutions

Tree 😆 DAG 🙁

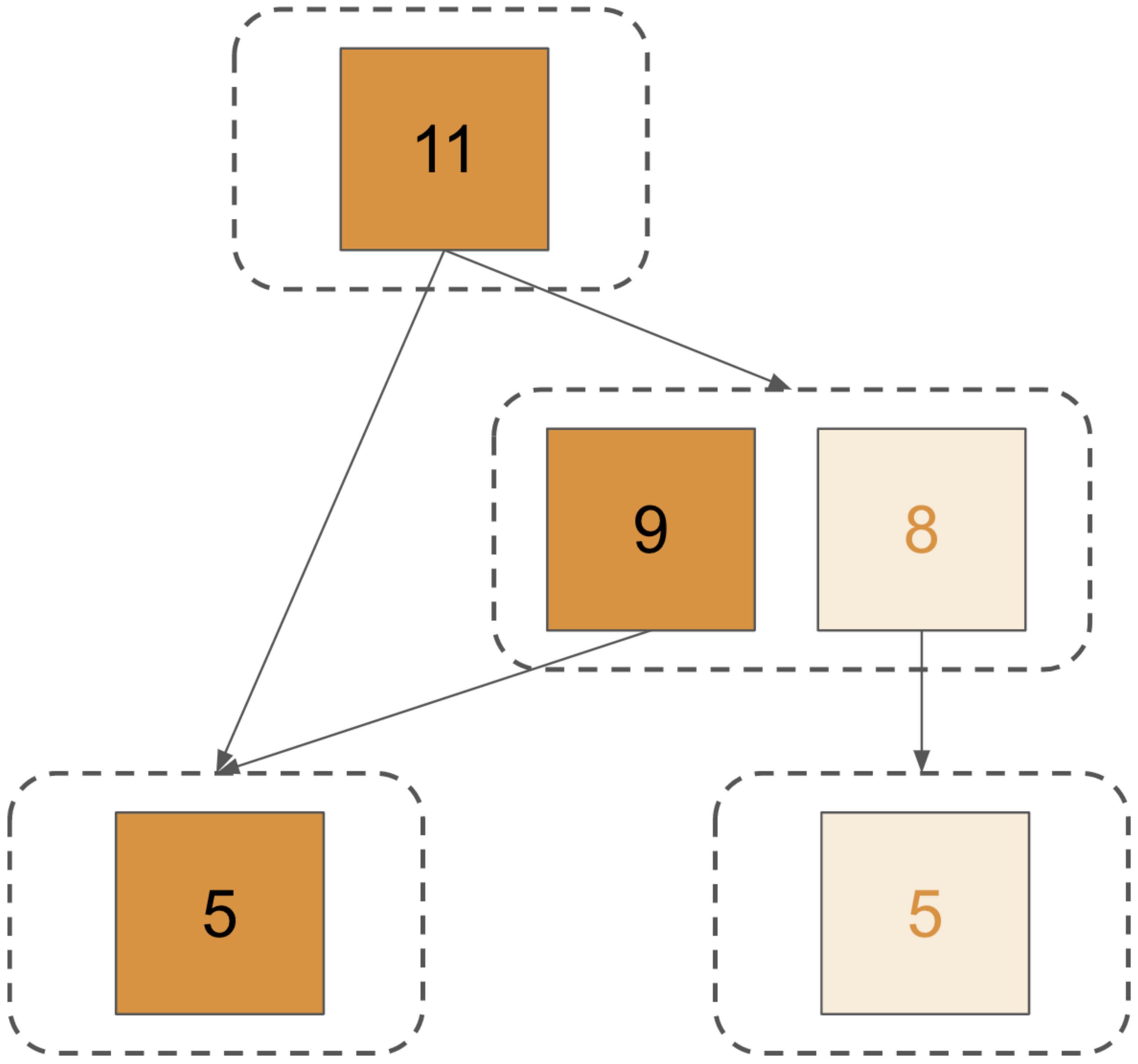


Yisu Remy Wang, Shana Hutchison, Jonathan Leang, Bill Howe, & Dan Suciu. (2020). SPORES: Sum-Product Optimization via Relational Equality Saturation for Large Scale Linear Algebra.
Yichen Yang, Phitchaya Mangpo Phothilimtha, Yisu Remy Wang, Max Willsey, Sudip Roy, & Jacques Pienaar. (2021). Equality Saturation for Tensor Graph Superoptimization.

Term Extraction



Extracted by Greedy
(29)



Optimal
(25)

Term Extraction

Alternative: Integer linear programming (ILP)

For each e-node n, create a **binary (0/1) variable** w_n (call them node variables)

Root Constraint: $\sum_n w_n \geq 1$

for all n in the root e-class

Children Constraint: $\sum_{n' \in C_i} w_{n'} \geq w_n$

for each child C_i of n

Objective: Minimize $\sum_n w_n \text{cost(n)}$

Yisu Remy Wang, Shana Hutchison, Jonathan Leang, Bill Howe, & Dan Suciu. (2020). SPORES: Sum-Product Optimization via Relational Equality Saturation for Large Scale Linear Algebra.
Yichen Yang, Phitchaya Mangpo Phothilimtha, Yisu Remy Wang, Max Willsey, Sudip Roy, & Jacques Pienaar. (2021). Equality Saturation for Tensor Graph Superoptimization.

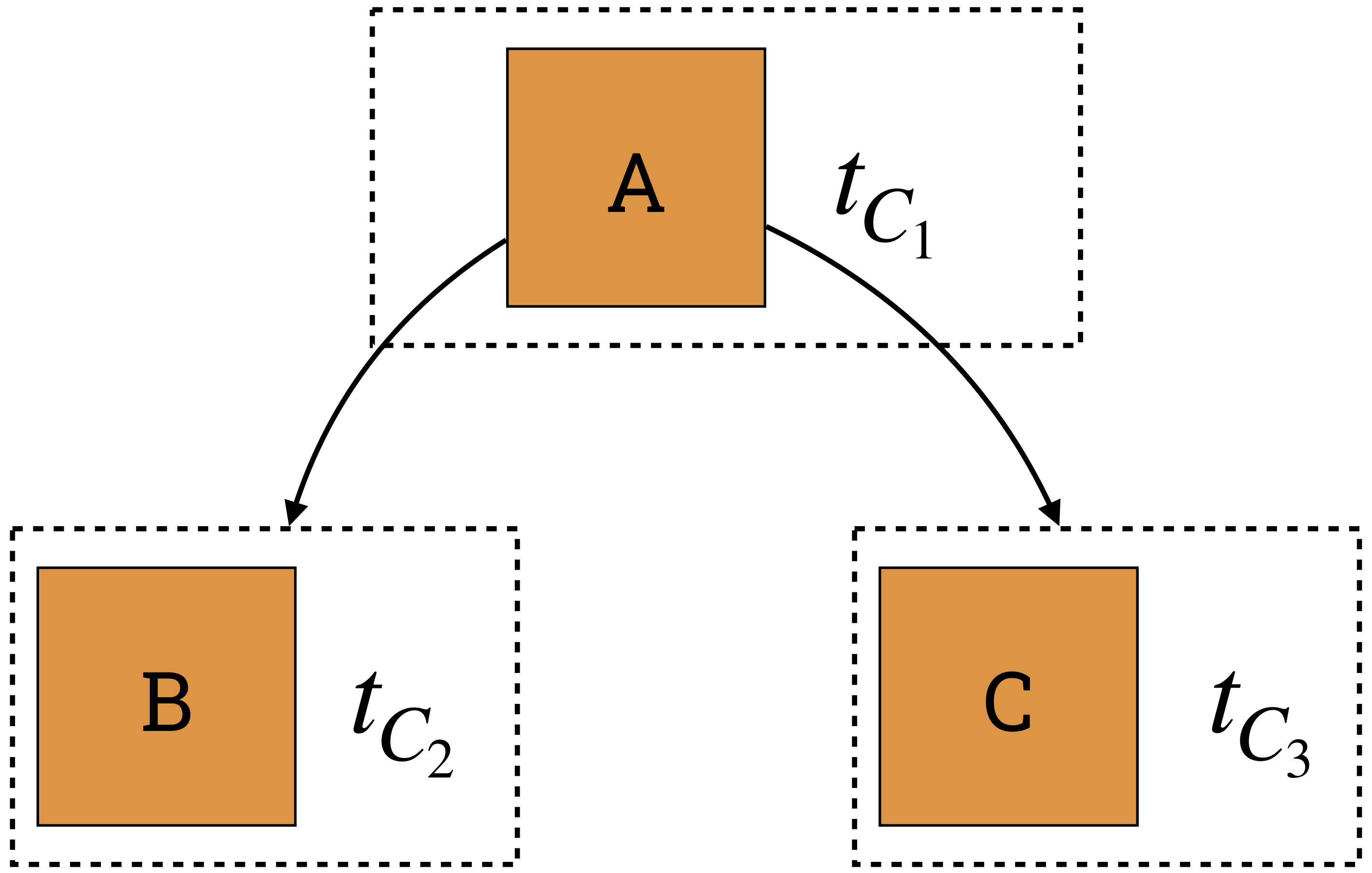
Term Extraction

Avoiding Cycles: Topological sorting (ILP-Topo)

For each e-class C, create an integral value t_C bounded by a sufficiently large value σ (e.g. 2x number of the e-classes)

“If we pick an e-node n, then the topological order of its e-class must be greater than those of all its children.”

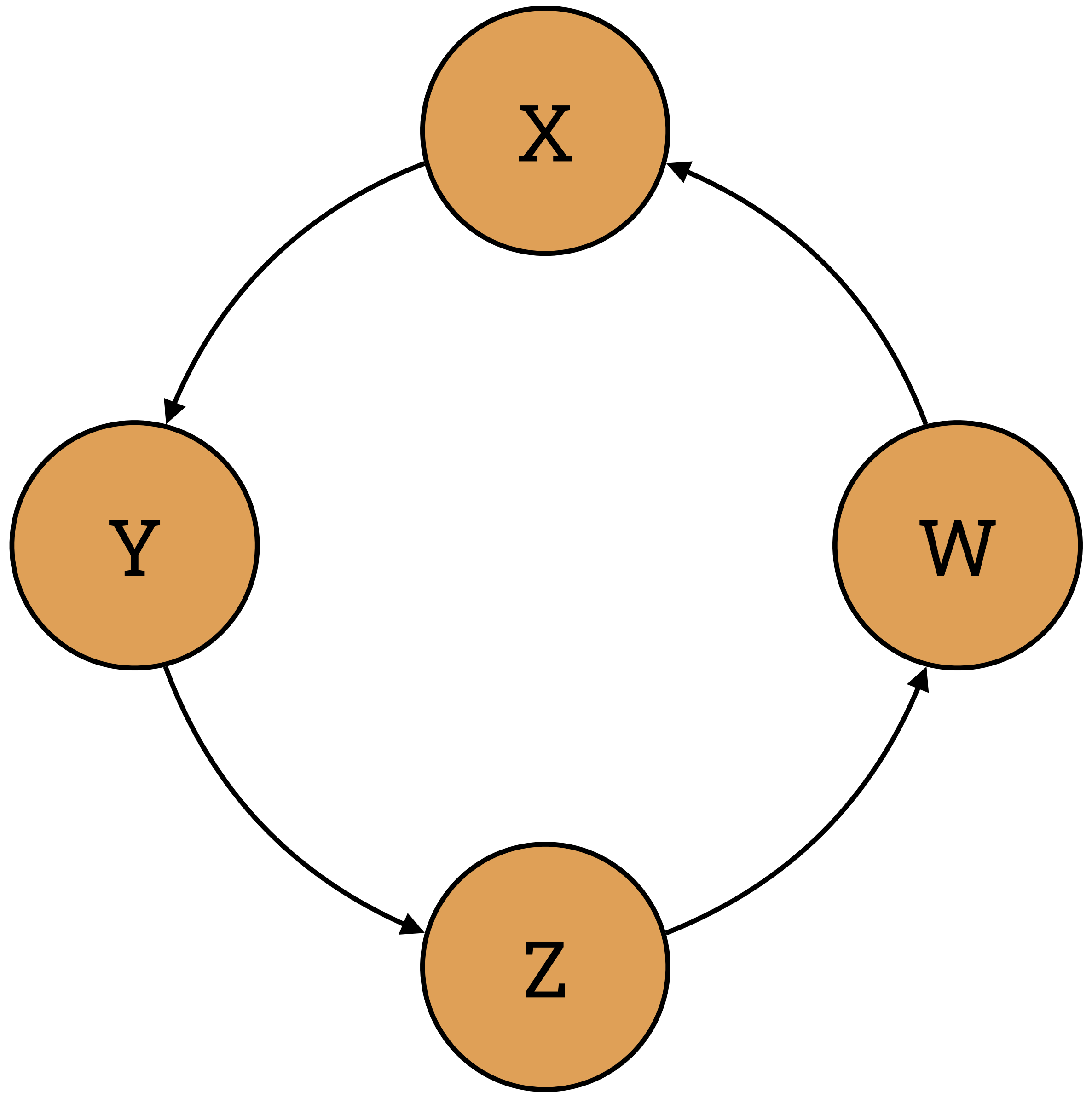
Too expensive!



$$t_{C_1} > t_{C_2}$$
$$t_{C_1} > t_{C_3}$$

Yisu Remy Wang, Shana Hutchison, Jonathan Leang, Bill Howe, & Dan Suciu. (2020). SPORES: Sum-Product Optimization via Relational Equality Saturation for Large Scale Linear Algebra.
Yichen Yang, Phitchaya Mangpo Phothilimtha, Yisu Remy Wang, Max Willsey, Sudip Roy, & Jacques Pienaar. (2021). Equality Saturation for Tensor Graph Superoptimization.

Acyclic Constraints

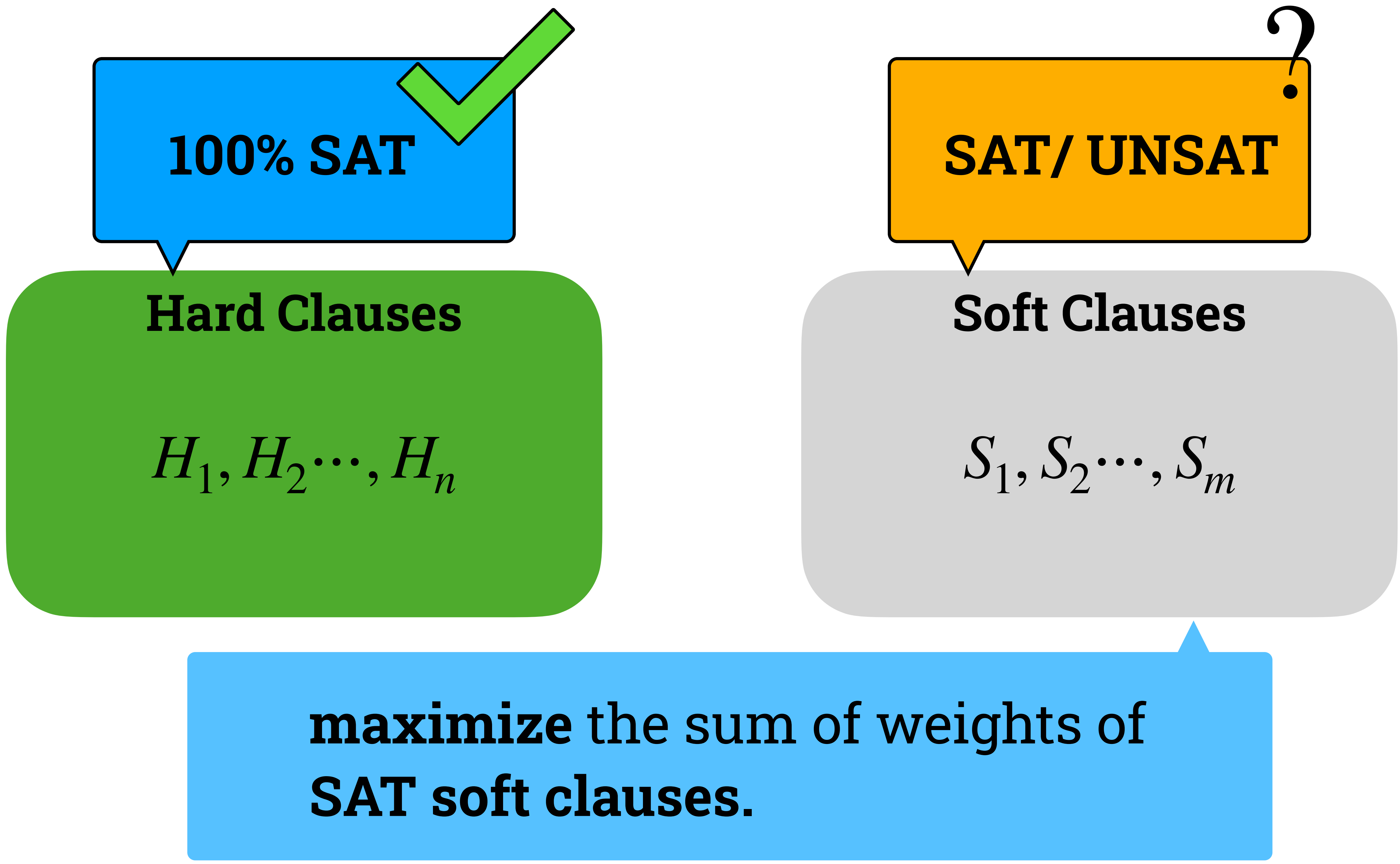


$$X + Y + Z + W \leq 3$$

$$\neg X \vee \neg Y \vee \neg Z \vee \neg W$$

Background: Weighted (Partial) MaxSAT

Weighted: clauses carry a positive weight.



Weighted MaxSAT Encoding

For each e-node n, create a **boolean variable** w_n (call them node variables)

Hard Clauses

Root Constraints
Children Constraints
Acyclic Constraints

Soft Clauses

$\neg w_n$

Weighted MaxSAT Encoding

Soft Clauses

$$\neg w_n$$

Soft constraints are simply $\neg w_n$ for each e-node n with weights cost(n).

Weighted MaxSAT Encoding

Hard Clauses

Root Constraint: $\bigvee w_{n_i}$

Children Constraints: $w_n \rightarrow \bigvee_{n' \in C_i} w_{n'}$

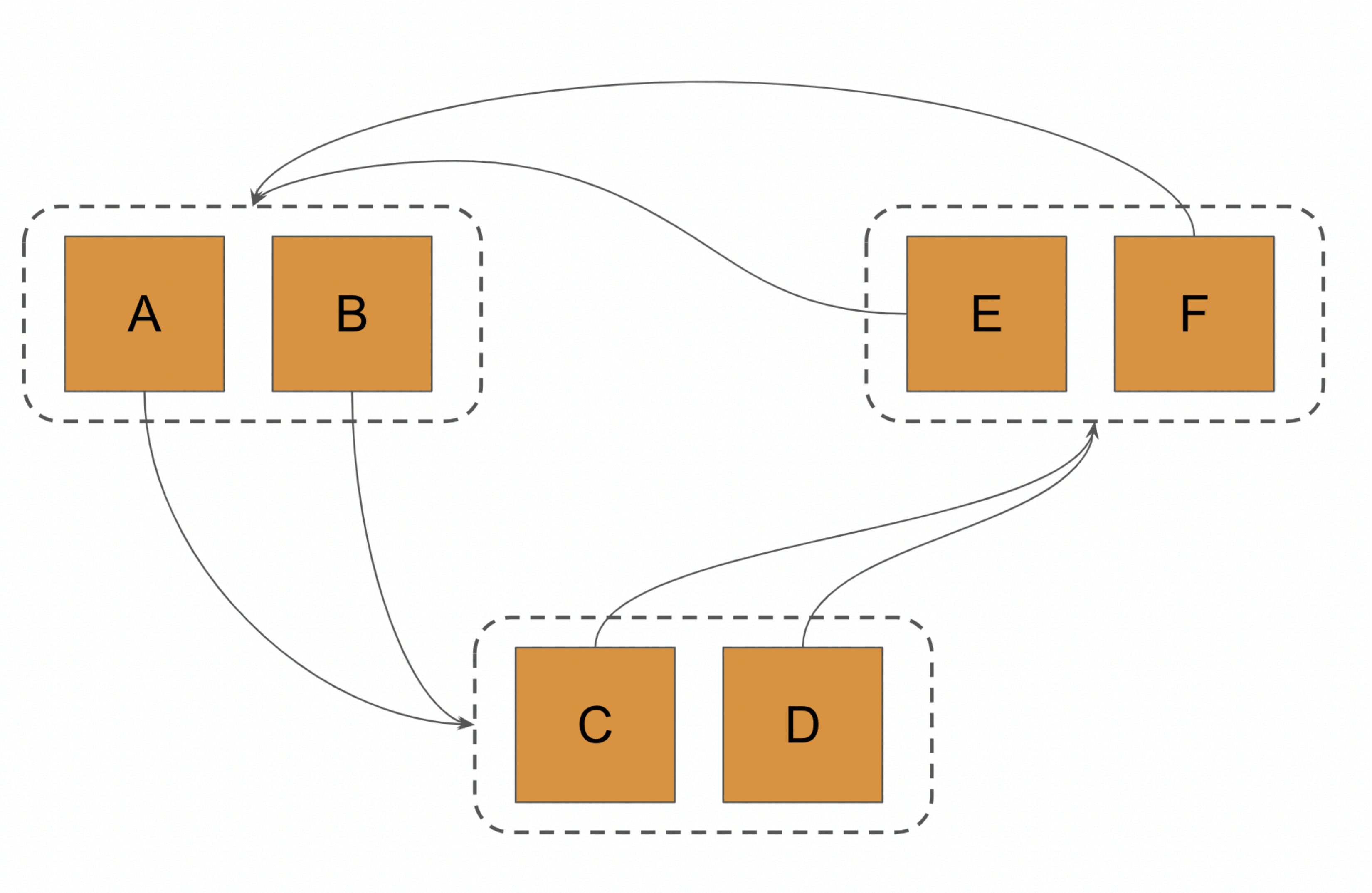
Acyclic constraints (naive): $\bigvee_{n \in \phi} \neg w_n$ (for all cycles ϕ)

Weighted MaxSAT Encoding

Reducing # of Acyclic Constraints

Acyclic constraints (naive): $\bigvee_{n \in \phi} \neg w_n$ (for all cycles ϕ)

The naïve encoding could yield exponentially many constraints for a single cycle of e-classes



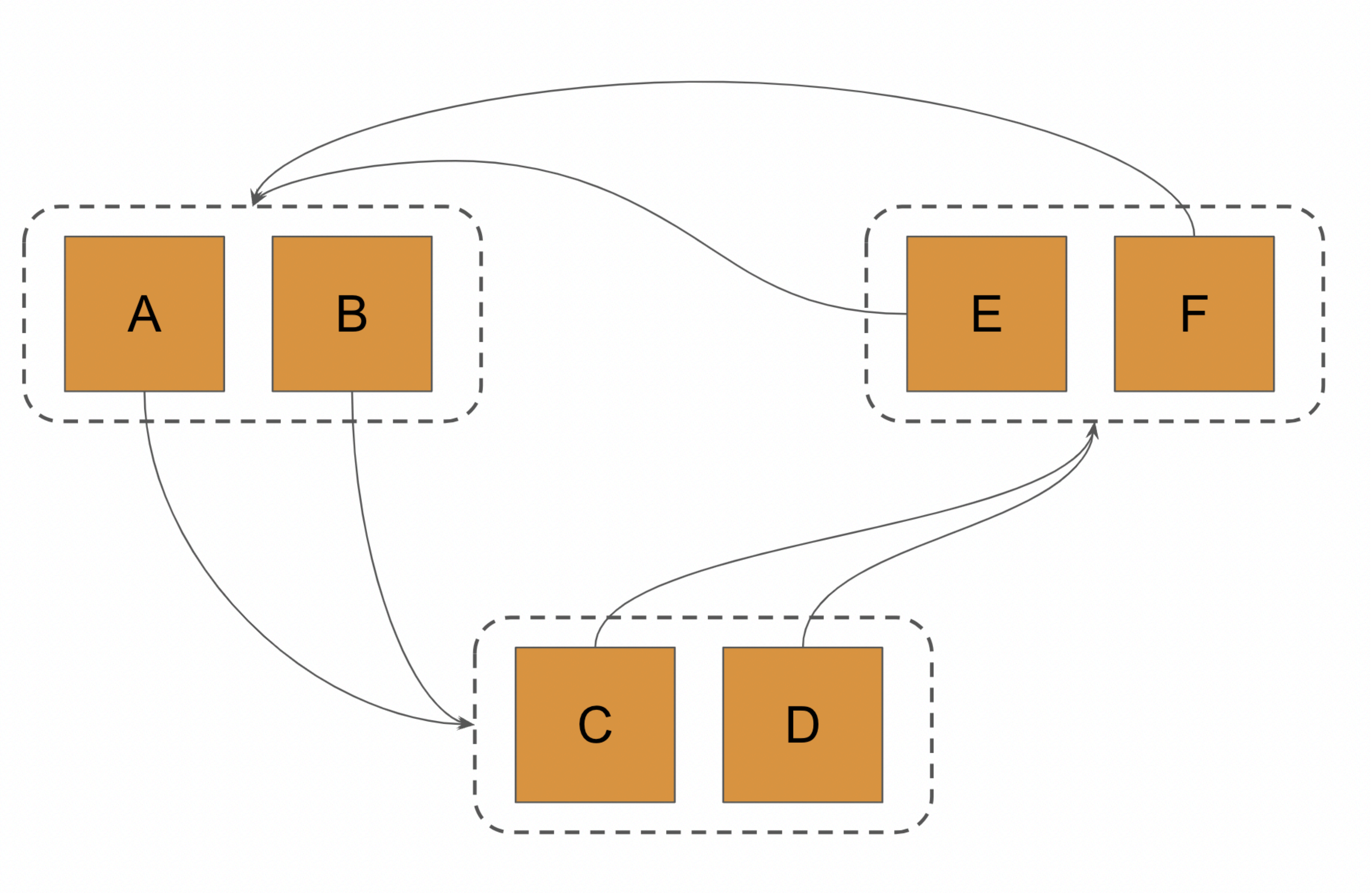
Weighted MaxSAT Encoding

Reducing # of Acyclic Constraints

Instead of encoding on cycles of e-nodes, we could instead work with **cycles of e-classes**

$$\bigvee_{C_i} \bigwedge_{n \in C_i \wedge in\_cycle(n)} \neg w_n$$

Then, perform Tseitin Transformation to get CNF



$$\text{Tseitin}\left((\neg w_A \wedge \neg w_B) \vee (\neg w_C \wedge \neg w_D) \vee (\neg w_E \wedge \neg w_F)\right)$$

$$\Leftrightarrow$$

$$x_{AB} \leftrightarrow (\neg w_A \wedge \neg w_B)$$

$$x_{CD} \leftrightarrow (\neg w_C \wedge \neg w_D)$$

$$x_{EF} \leftrightarrow (\neg w_E \wedge \neg w_F)$$

$$x_{AB} \vee x_{CD} \vee x_{EF}$$

ILP Encoding

Replacing the topological ordering constraints with acyclic constraints.

$$x_{C_i} \leftrightarrow \bigwedge \neg w_j$$

If direction: $(1 - x_{C_i}) + (1 - w_j) \geq 1$

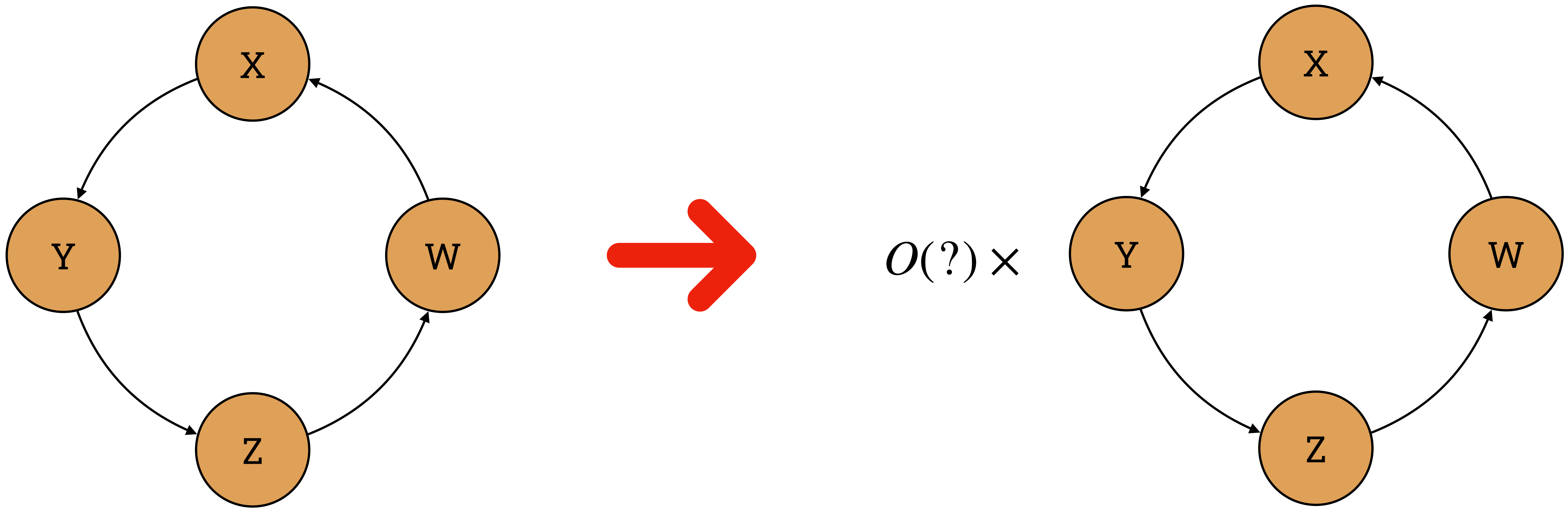
Only-if direction: $x_{C_i} + \sum w_j \geq 1$

Following Tseitin Transformation: $\sum x_{C_i} \geq 1$

Call this encoding as **ILP-ACyc**

WPMAXSAT and ILP-ACyc

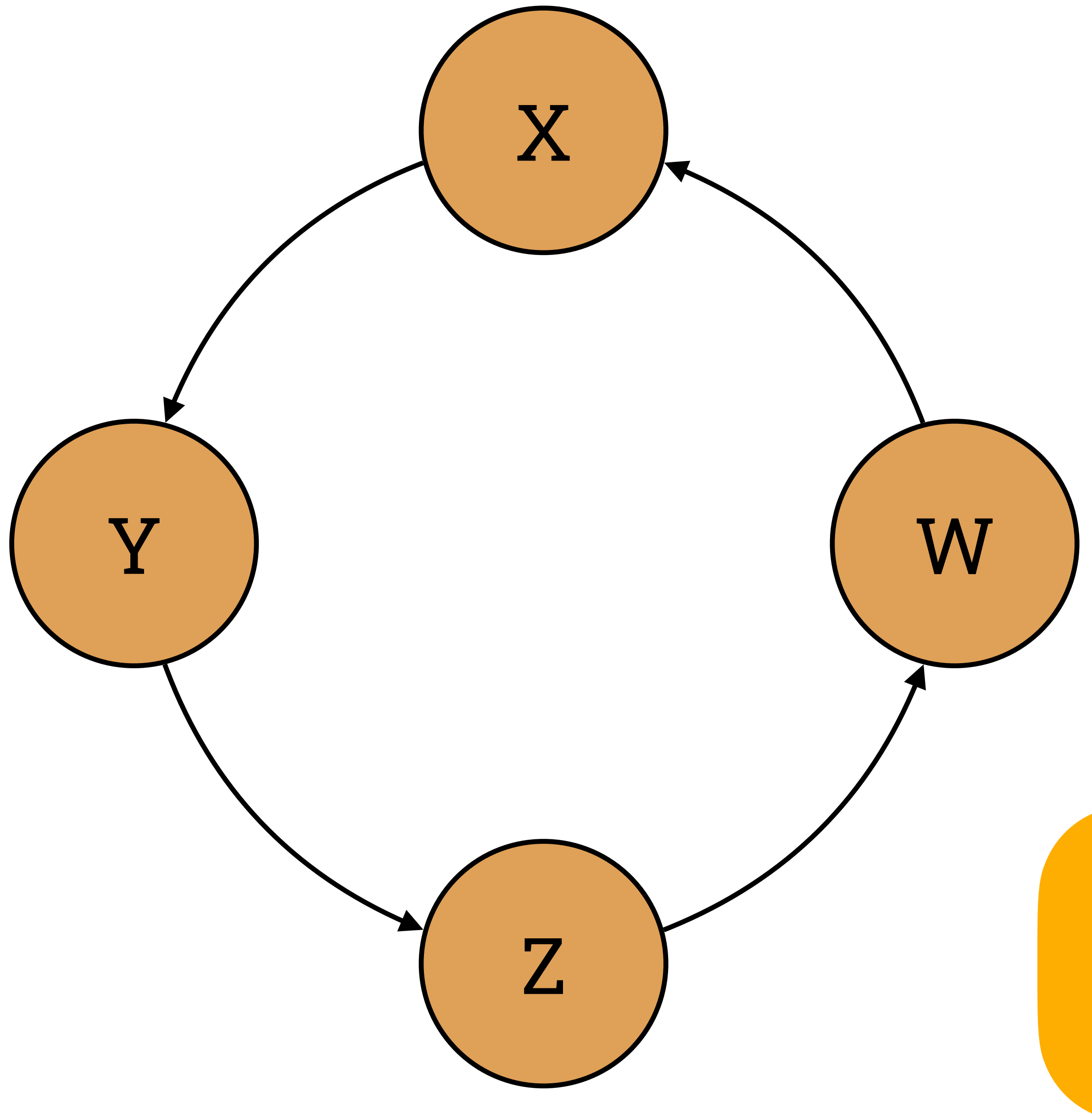
Remaining Issue: # of cycles of e-classes?



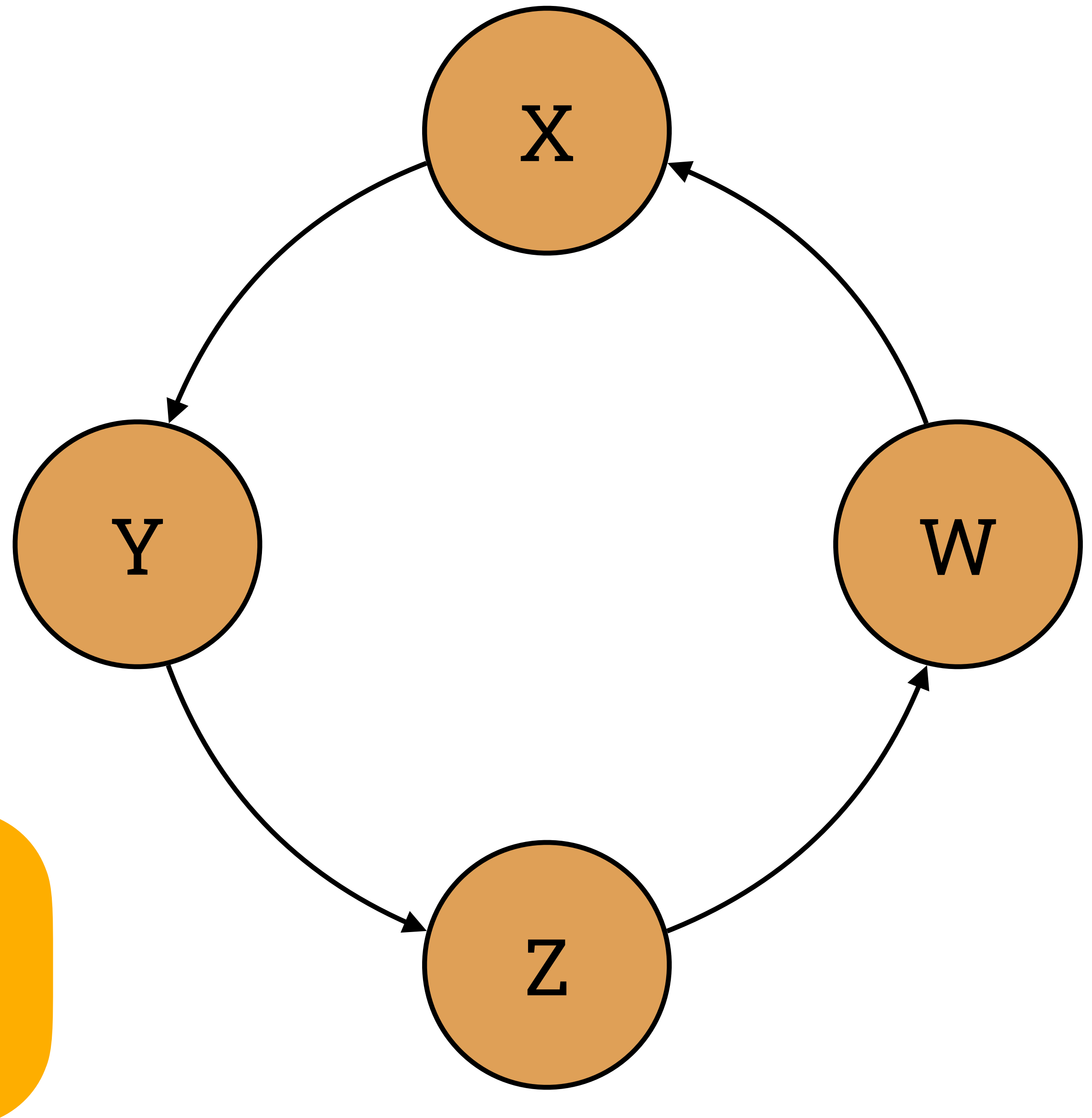
$O(n)$ constraints per e-class cycle

WPMAXSAT and ILP-ACyc

Remaining Issue: # of cycles of e-classes?



$O(?) \times$



Our assumption: worst case is <u>unlikely</u>

$O(n)$ constraints per e-class cycle

Worst case...$O(2^{|C|})$

Experiments

Evaluation setup

Workload:
Extracting optimal terms from saturated e-graphs from Glenside[1]
Tensor programs are obtained from ResNet-18/50, MobileNet, ResMLP and EfficientNet.

Rewrite Rules:
Im2Col: image-to-column transformations
Im2Col + SIMPL: Im2Col plus a set of simplification rewrites, including

- Operator collapsing (transpose, reshape, access, etc.)
- Operator reordering

Configurations:
5-second timeout for equality saturation
5-minute timeout for term extraction (including time of constructing constraints)

1: Gus Henry Smith, Andrew Liu, Steven Lyubomirsky, Scott Davidson, Joseph McMahan, Michael Taylor, Luis Ceze, & Zachary Tatlock (2021). Pure tensor program rewriting via access patterns (representation pearl). In *Proceedings of the 5th ACM SIGPLAN International Symposium on Machine Programming*. ACM.

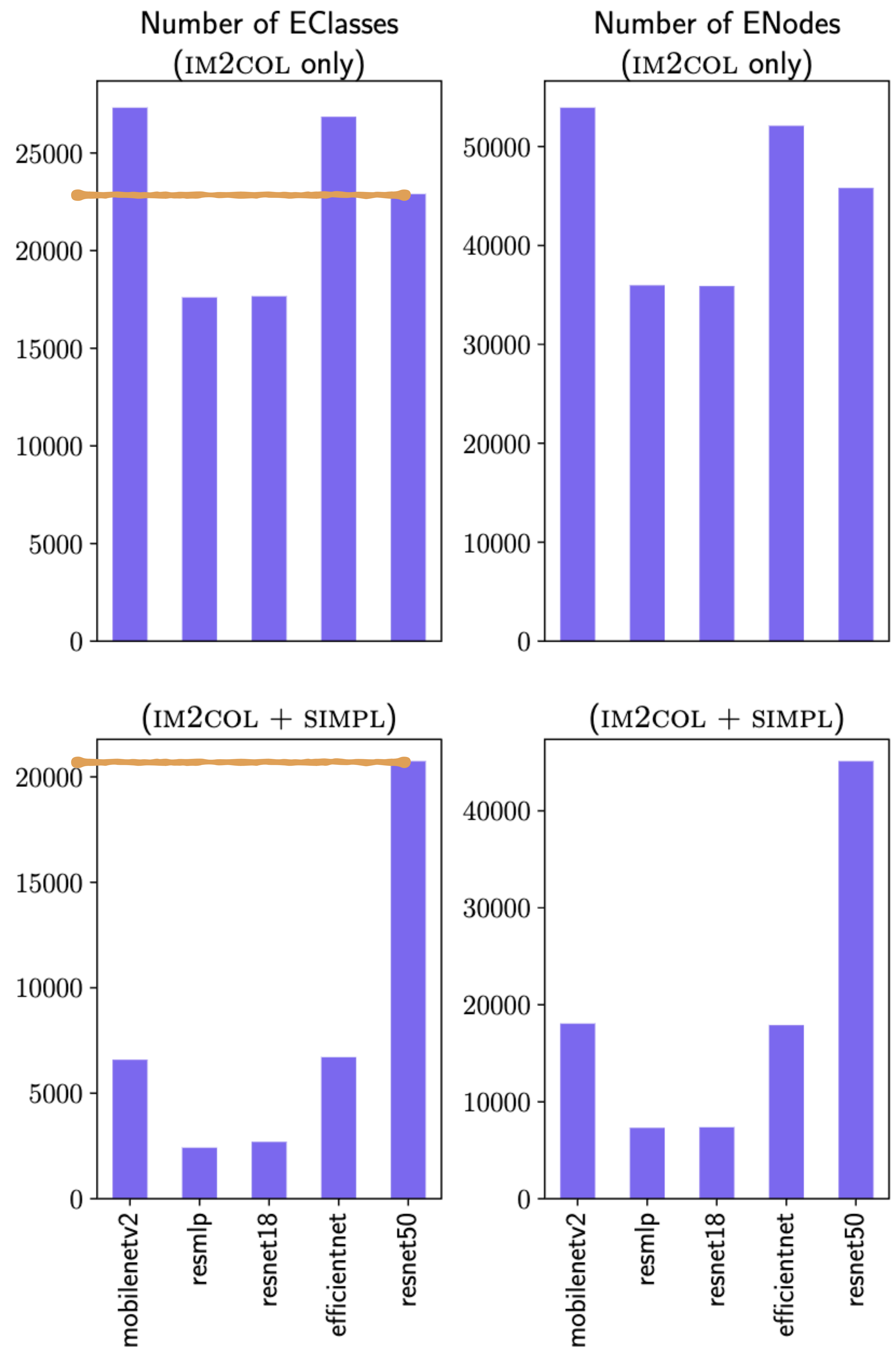
Results

E-graph statistics

	MobileNet	ResMLP	ResNet-18	EfficientNet	ResNet-50
IM2COL	17266	15819	14754	21016	16305
IM2COL + SIMPL	17320	4247	4466	10978	20294

Table 1. Class cycle count after equality saturation

Good News!

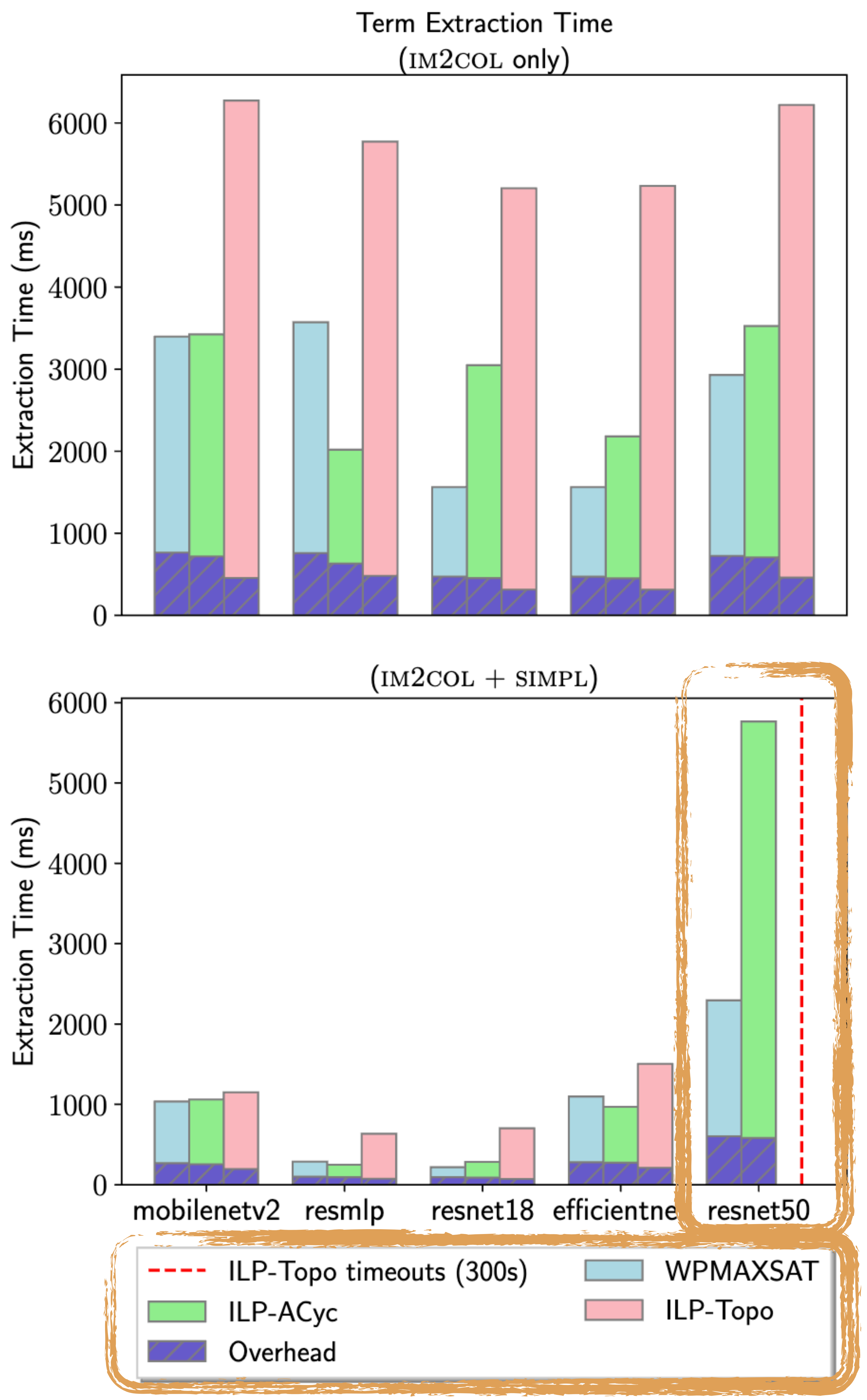


Results

Constraint building + term extraction time

Topological Sorting

Acyclic Constraints



Improving Term Extraction with Acyclic Constraints

Mike He, Haichen Dong, Sharad Malik and Aarti Gupta

TY!

PRINCETON UNIVERSITY

Discussion

LP Relaxation of ILP-Topo

Relaxation is not trivial:

"Weight vanishing"
Recall Children constraints:

$$-w_n + \sum_{n' \in C_i} w_{n'} \geq 0$$

w_n could be distributed over e-nodes in its children e-classes.