

Boosting Constrained Horn Solving by Unsat Core Learning

Parosh Aziz Abdulla¹, Chencheng Liang¹, Philipp Rümmer^{1,2}

¹Uppsala University, Sweden

²University of Regensburg, Germany

January 15, VMCAI 2024



UPPSALA
UNIVERSITET



Universität Regensburg

Background

- Solving Constrained Horn Clauses (CHCs)

Background

- Solving Constrained Horn Clauses (CHCs)

A CHC is a formula in the format

$$\forall V. L[X] \leftarrow L_1[X_1] \wedge \dots \wedge L_n[X_n] \wedge \varphi$$

Where

V are variables,

X_i are terms over V ,

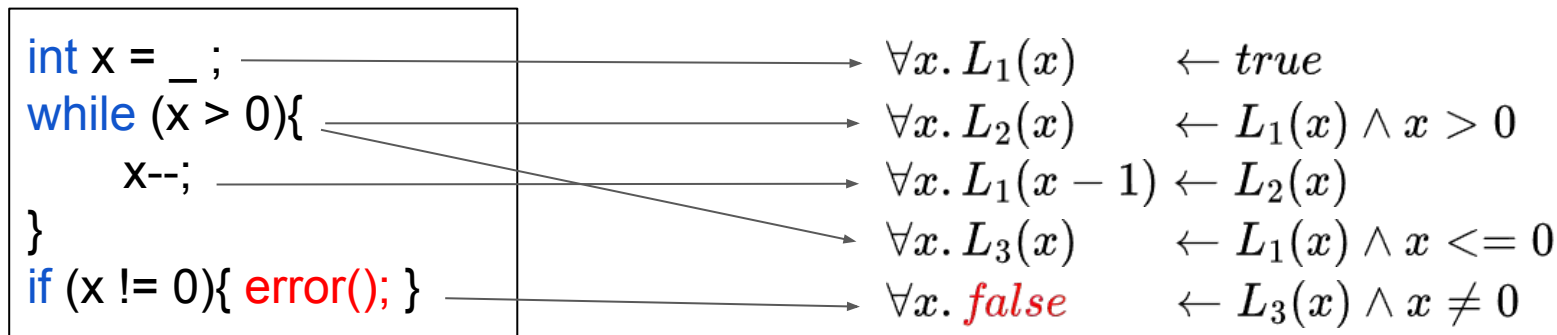
$L, L_1, \dots, L_n$ are n -ary relation symbols,

$L_i[X_i]$ is an atom of relation symbol to the terms,

φ is a constraint in the background theory T .

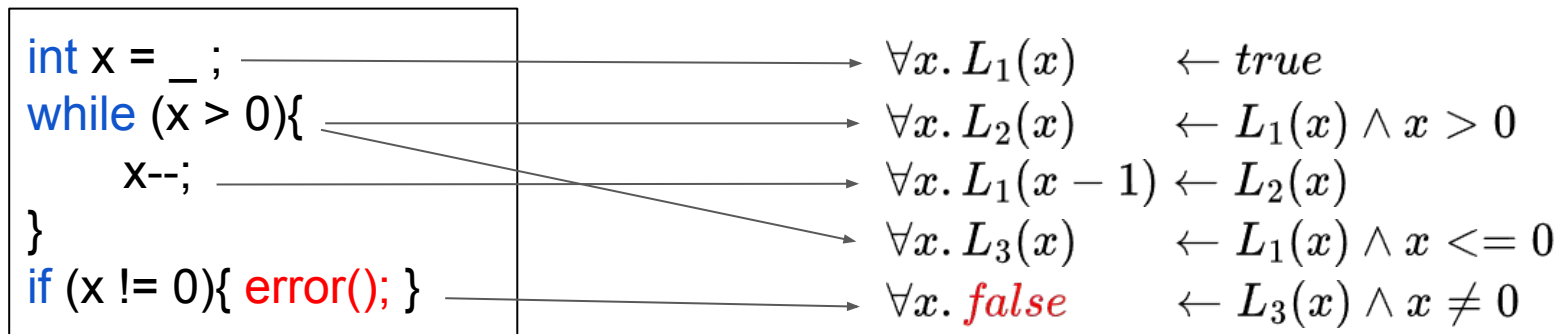
A Motivating Example

- A program verification problem encoded as CHCs



A Motivating Example

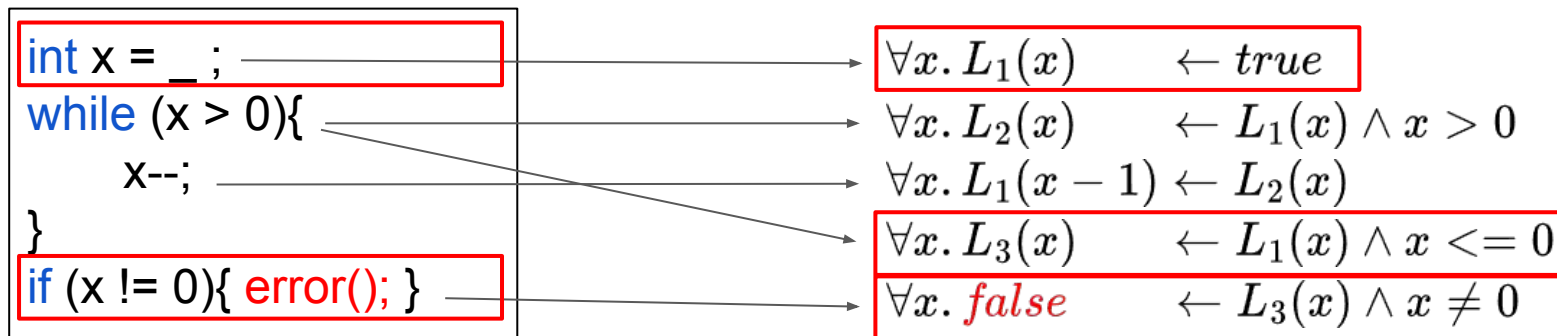
- A program verification problem encoded as CHCs



If any subset of the set of CHCs is UNSAT, then the entire set of CHCs is also UNSAT.

A Motivating Example

- A program verification problem to CHCs



If any subset of the set of CHCs is UNSAT, then the entire set of CHCs is also UNSAT.

Minimal Unsatisfiable Subsets (MUSes) of CHCs

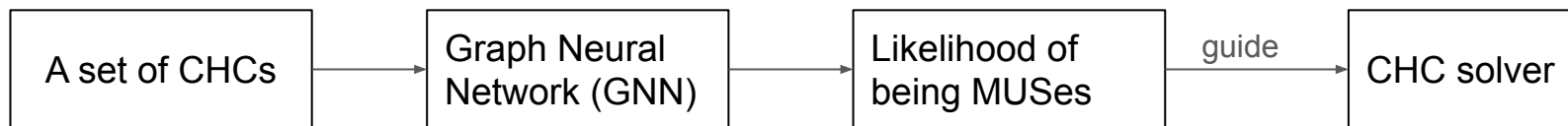
- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$

Minimal Unsatisfiable Subsets (MUSes) of CHCs

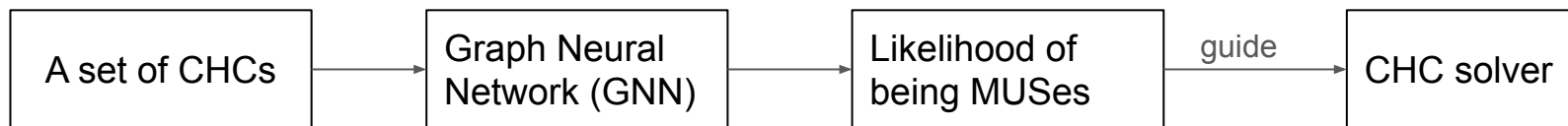
- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$

{[1], [4], [5]} is the only MUS

Deep Learning-Based Framework

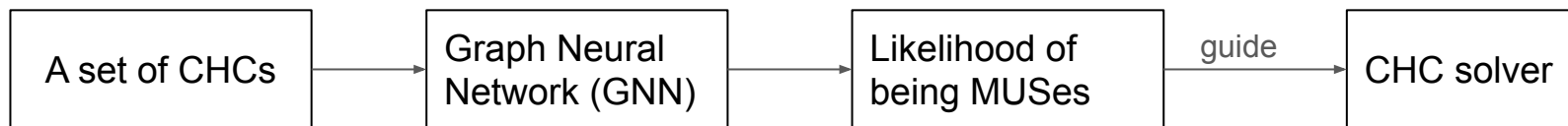


Deep Learning-Based Framework



- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$

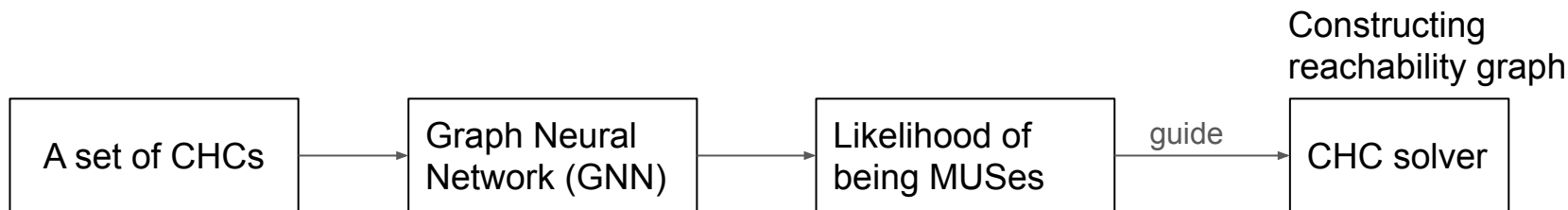
Deep Learning-Based Framework



Score

0.8	[1]	$L_1(x)$	$\leftarrow true$
0.2	[2]	$L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3]	$L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4]	$L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5]	$false$	$\leftarrow L_3(x) \wedge x \neq 0$

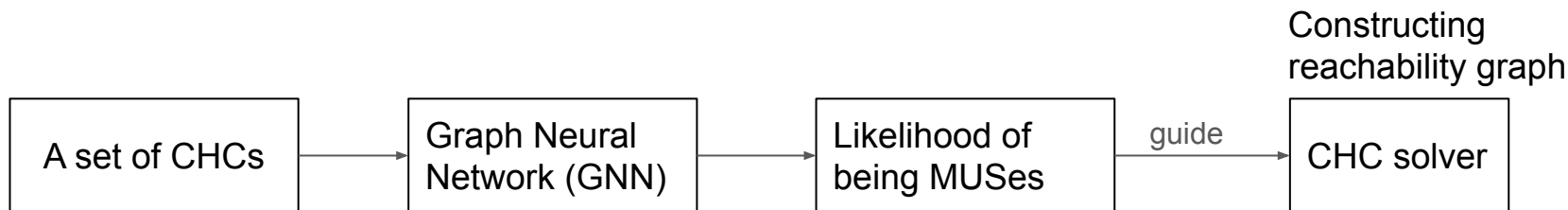
Deep Learning-Based Framework



Score

0.8	[1] $L_1(x)$	$\leftarrow true$
0.2	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

Deep Learning-Based Framework

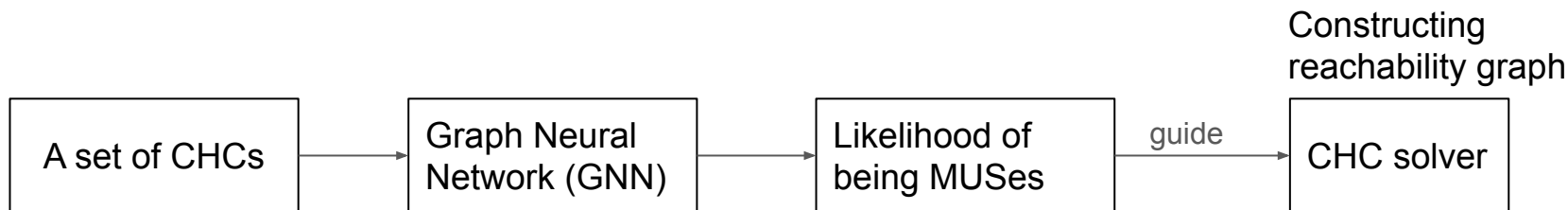


Score

0.8	[1]	$L_1(x)$	$\leftarrow true$
0.2	[2]	$L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3]	$L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4]	$L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5]	$false$	$\leftarrow L_3(x) \wedge x \neq 0$

The order of CHCs is important

Deep Learning-Based Framework



Score

0.8	[1]	$L_1(x)$	$\leftarrow true$
0.2	[2]	$L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3]	$L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4]	$L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5]	$false$	$\leftarrow L_3(x) \wedge x \neq 0$

The order of CHCs is important

Prioritized CHCs:

[1], [4], [5], [2], [3]

Deep Learning-Based Framework (extract training data)



Deep Learning-Based Framework (extract training data)



Label

Clauses

1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

Deep Learning-Based Framework (extract training data)



Label

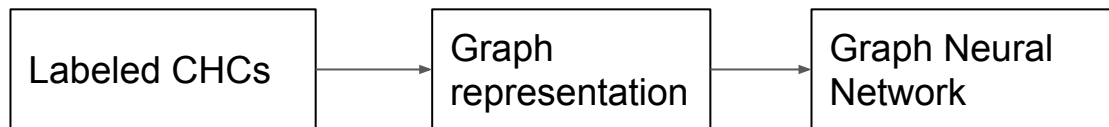
Clauses

1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

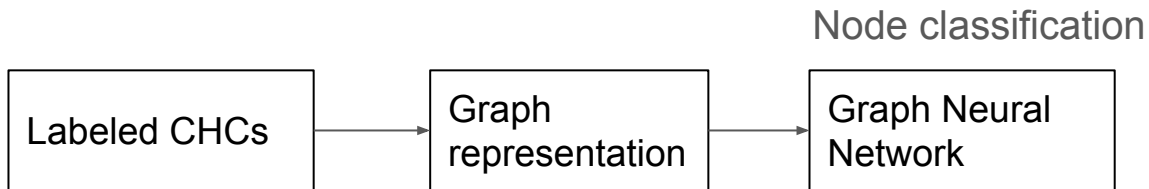
When there are multiple MUSes

- Union
- Intersection
- Single

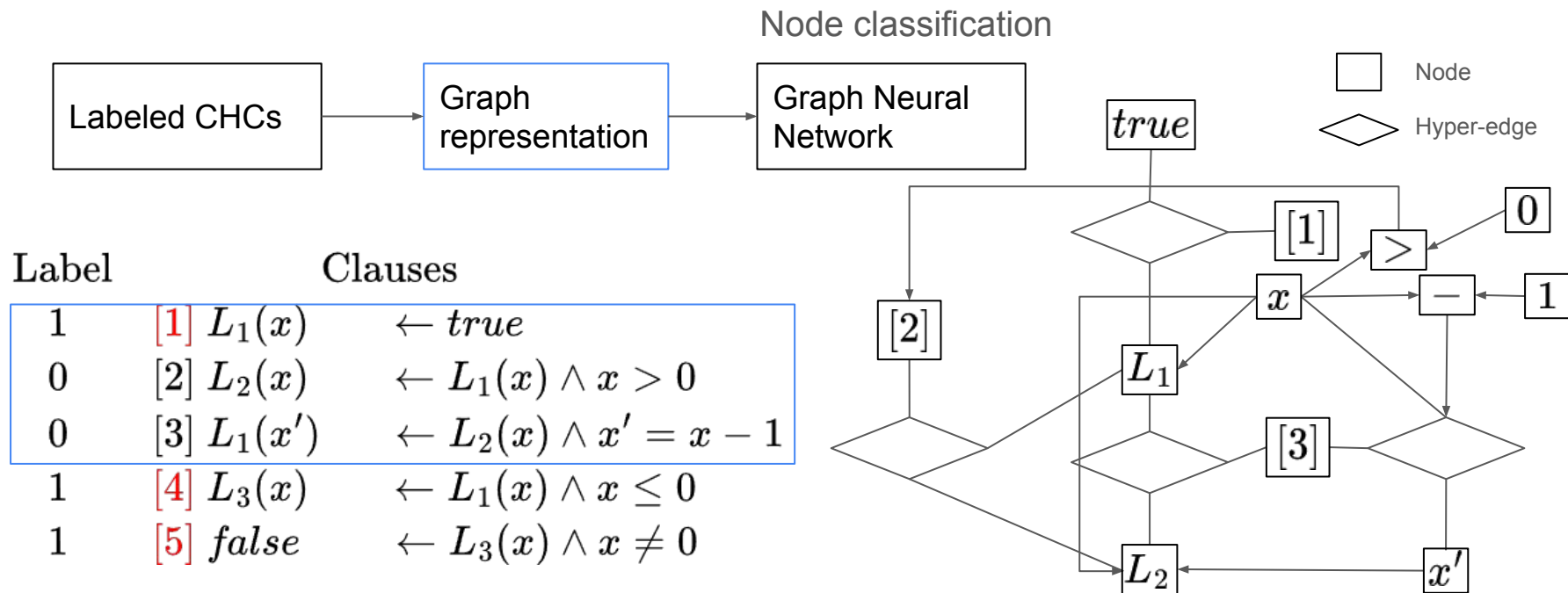
Deep Learning-Based Framework (train a model)



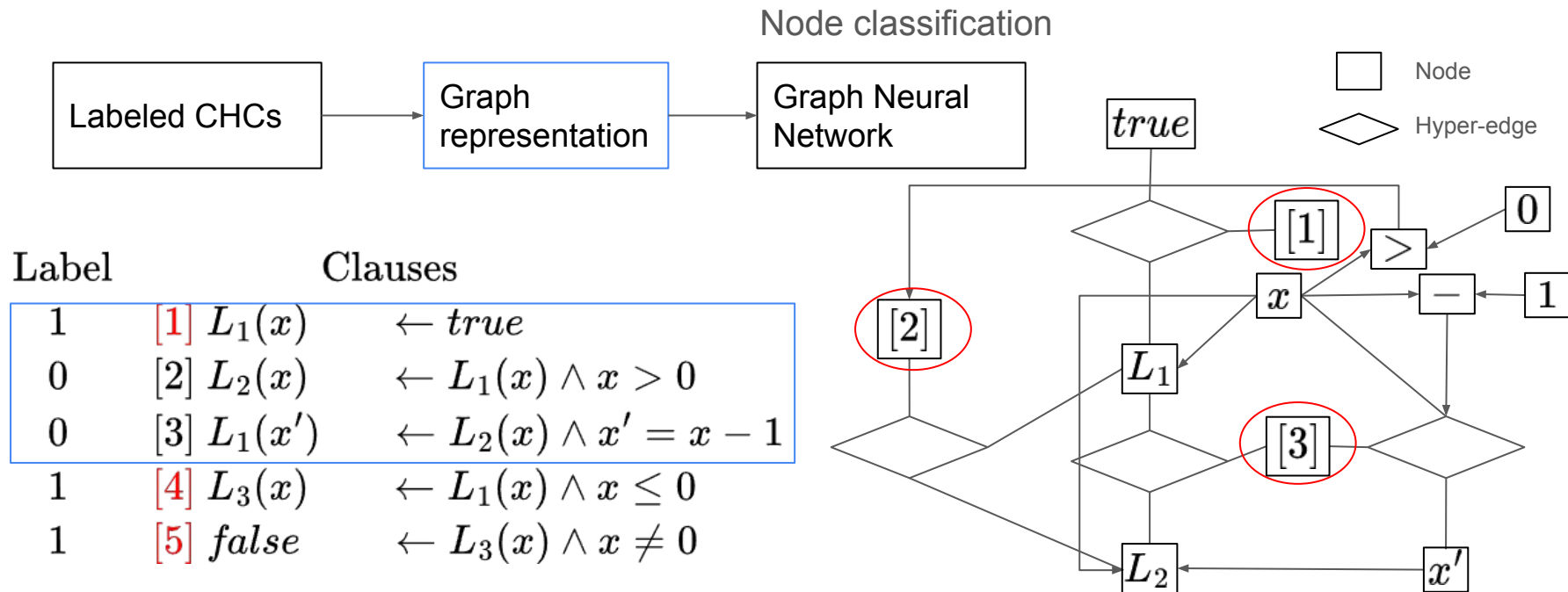
Deep Learning-Based Framework (train a model)



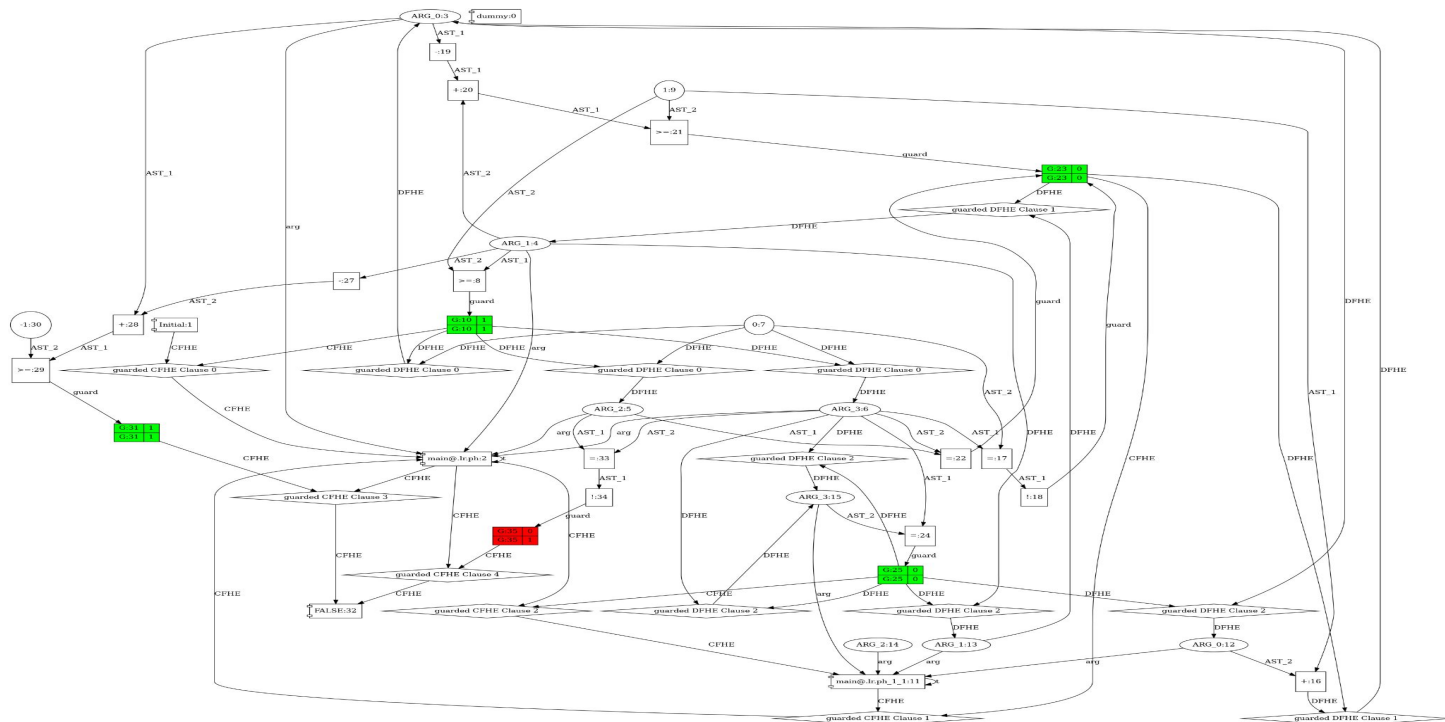
Deep Learning-Based Framework (train a model)



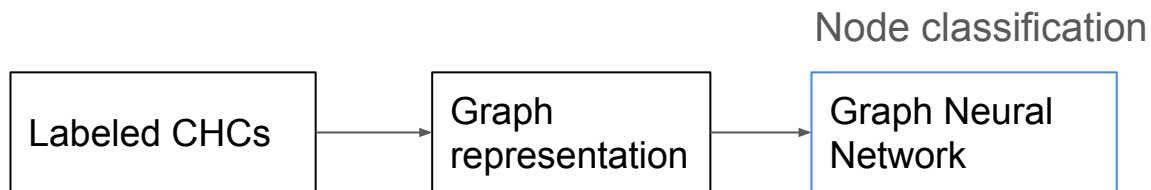
Deep Learning-Based Framework (train a model)



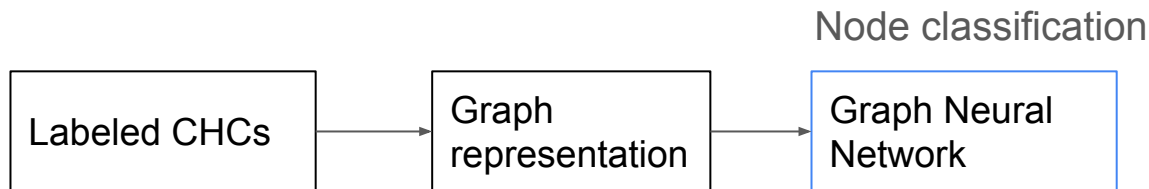
Represent CHCs by graphs



Deep Learning-Based Framework (train a model)



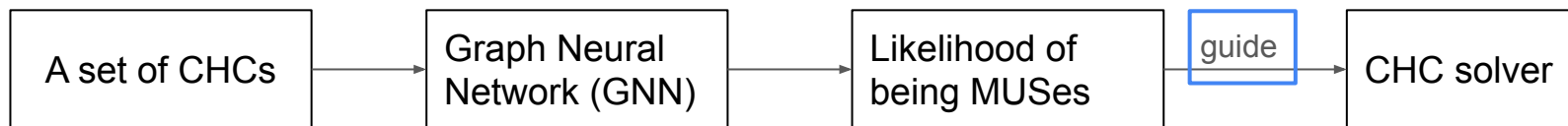
Deep Learning-Based Framework (train a model)



- Relational Hypergraph Neural Network [1]
 - Can handle different types of hyperedges

[1] Chencheng Liang, Philipp Rümmer, and Marc Brockschmidt.
Exploring Representation of Horn Clauses using GNNs

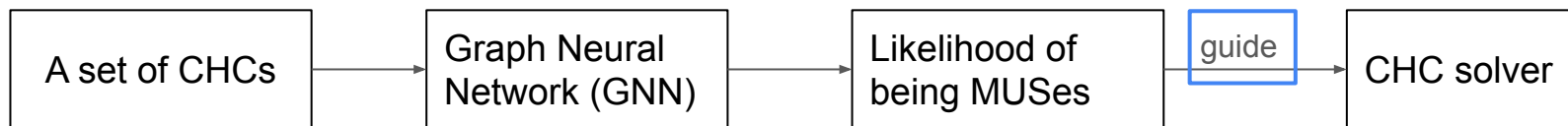
Deep Learning-Based Framework



Score

0.8	[1] $L_1(x)$	$\leftarrow true$
0.2	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

Deep Learning-Based Framework



- Use scores alone
- Combine with original prioritizing scores
 - Add/subtract normalized or ranked scores with coefficient
 - Randomly shifting between MUS and original score

Experimental results

- Benchmarks from CHC-COMP

Linear LIA problems				
8705				
Benchmarks for training			Holdout set	
7834 (90%)			871 (10%)	
UNSAT	SAT	T/O	Eval.	N/A
1585	4004	2245	383	488
Train	Valid	N/A		
782	87	716		

Experimental results

- Benchmarks from CHC-COMP

Linear LIA problems					Non-linear LIA problems				
8705					8425				
Benchmarks for training			Holdout set		Benchmarks for training			Holdout set	
7834 (90%)			871 (10%)		7579 (90%)			846 (10%)	
UNSAT	SAT	T/O	Eval.	N/A	UNSAT	SAT	T/O	Eval.	N/A
1585	4004	2245	383	488	3315	4010	254	488	358
Train	Valid	N/A			Train	Valid	N/A		
782	87	716			1617	180	1518		

Experimental results

- Algorithms of CHC solver
 - Counterexample-guided abstraction refinement (CEGAR)
 - Symbolic execution (SymEx)

Experimental results (Improved percentage)

Benchmark Algorithm	MUS data set (best count)	Best ranking function (improvement in %)						
		Number of Solved Problems			Average Time			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Linear CEGAR	Union (0)	R-Plus (1.4%)	R-Plus (2.4%)	R-Minus (1.0%)	R-Plus (1.3%)	S-Plus (19.1%)	S-Minus (46.5%)	Rank (31.1%)
	Single (3)	Rank (3.6%)	R-Plus (4.0%)	Rank (8.2%)	R-Plus (1.9%)	S-Plus (26.6%)	R-Minus (57.9%)	Rank (36.3%)
	Intersection (4)	R-Plus (4.1%)	S-Plus (0.8%)	R-Plus (9.3%)	R-Plus (3.1%)	S-Plus (27.6%)	R-Minus (45.0%)	S-Plus (0.0%)
Linear SymEx	Union (4)	Two-Q (1.0%)	S-Plus* (0.0%)	Random (2.0%)	Two-Q (0.9%)	R-Minus (12.7%)	R-Minus (30.2%)	S-Plus (26.5%)
	Single (3)	S-Minus* (0.5%)	S-Plus* (0.0%)	Random (2.0%)	Random (0.8%)	S-Plus (12.9%)	Random (28.4%)	S-Plus (17.6%)
	Intersection (5)	S-Plus* (1.0%)	S-Plus* (0.0%)	S-Plus* (2.0%)	S-Plus (1.3%)	Score (9.5%)	Random (28.4%)	R-Plus (35.8%)

Experimental results (Improved percentage)

Benchmark Algorithm	MUS data set (best count)	Best ranking function (improvement in %)						
		Number of Solved Problems			Average Time			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Linear CEGAR	Union (0)	R-Plus (1.4%)	R-Plus (2.4%)	R-Minus (1.0%)	R-Plus (1.3%)	S-Plus (19.1%)	S-Minus (46.5%)	Rank (31.1%)
	Single (3)	Rank (3.6%)	R-Plus (4.0%)	Rank (8.2%)	R-Plus (1.9%)	S-Plus (26.6%)	R-Minus (57.9%)	Rank (36.3%)
	Intersection (4)	R-Plus (4.1%)	S-Plus (0.8%)	R-Plus (9.3%)	R-Plus (3.1%)	S-Plus (27.6%)	R-Minus (45.0%)	S-Plus (0.0%)
Linear SymEx	Union (4)	Two-Q (1.0%)	S-Plus* (0.0%)	Random (2.0%)	Two-Q (0.9%)	R-Minus (12.7%)	R-Minus (30.2%)	S-Plus (26.5%)
	Single (3)	S-Minus* (0.5%)	S-Plus* (0.0%)	Random (2.0%)	Random (0.8%)	S-Plus (12.9%)	Random (28.4%)	S-Plus (17.6%)
	Intersection (5)	S-Plus* (1.0%)	S-Plus* (0.0%)	S-Plus* (2.0%)	S-Plus (1.3%)	Score (9.5%)	Random (28.4%)	R-Plus (35.8%)

Experimental results (Improved percentage)

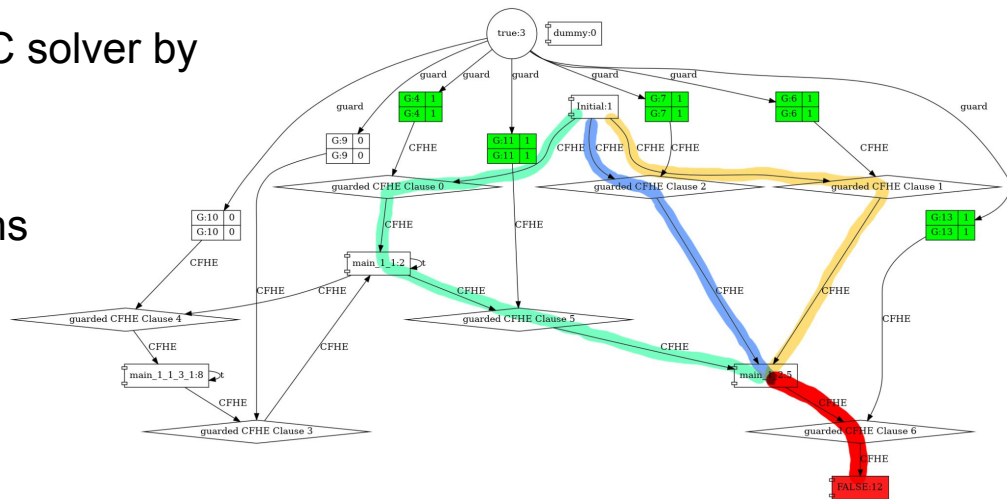
Benchmark Algorithm	MUS data set (best count)	Best ranking function (improvement in %)						
		Number of Solved Problems			Average Time			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Non-Linear CEGAR	Union (7)	S-Plus (0.5%)	S-Plus (0.8%)	S-Plus* (0.0%)	S-Plus (7.1%)	R-Minus (20.8%)	Rank (53.5%)	S-Plus (19.4%)
	Single (1)	R-Plus (0.2%)	R-Plus (0.4%)	R-Plus* (0.0%)	R-Plus (6.6%)	S-Plus (18.4%)	R-Minus (52.8%)	R-Minus (14.2%)
	Intersection (1)	R-Plus* (0.0%)	S-Plus (0.5%)	S-Plus* (0.0%)	R-Plus (5.9%)	R-Plus (20.3%)	Rank (45.8%)	S-Plus (16.8%)
Non-Linear SymEx	Union (6)	Two-Q (6.1%)	S-Minus* (1.6%)	Random (12.3%)	Two-Q (13.3%)	R-Minus (7.3%)	Score (5.1%)	R-Plus (19.9%)
	Single (3)	Two-Q (6.1%)	Score (1.6%)	Two-Q (12.9%)	Two-Q (12.4%)	Rank (-2.2%)	R-Minus (0.2%)	Two-Q (11.2%)
	Intersection (3)	Two-Q (6.1%)	S-Plus (1.6%)	Two-Q (12.9%)	Two-Q (12.7%)	S-Minus (0.6%)	Two-Q (1.7%)	S-Plus (5.4%)

Experimental results (Improved percentage)

Benchmark Algorithm	MUS data set (best count)	Best ranking function (improvement in %)						
		Number of Solved Problems			Average Time			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Non-Linear CEGAR	Union (7)	S-Plus (0.5%)	S-Plus (0.8%)	S-Plus* (0.0%)	S-Plus (7.1%)	R-Minus (20.8%)	Rank (53.5%)	S-Plus (19.4%)
	Single (1)	R-Plus (0.2%)	R-Plus (0.4%)	R-Plus* (0.0%)	R-Plus (6.6%)	S-Plus (18.4%)	R-Minus (52.8%)	R-Minus (14.2%)
	Intersection (1)	R-Plus* (0.0%)	S-Plus (0.5%)	S-Plus* (0.0%)	R-Plus (5.9%)	R-Plus (20.3%)	Rank (45.8%)	S-Plus (16.8%)
Non-Linear SymEx	Union (6)	Two-Q (6.1%)	S-Minus* (1.6%)	Random (12.3%)	Two-Q (13.3%)	R-Minus (7.3%)	Score (5.1%)	R-Plus (19.9%)
	Single (3)	Two-Q (6.1%)	Score (1.6%)	Two-Q (12.9%)	Two-Q (12.4%)	Rank (-2.2%)	R-Minus (0.2%)	Two-Q (11.2%)
	Intersection (3)	Two-Q (6.1%)	S-Plus (1.6%)	Two-Q (12.9%)	Two-Q (12.7%)	S-Minus (0.6%)	Two-Q (1.7%)	S-Plus (5.4%)

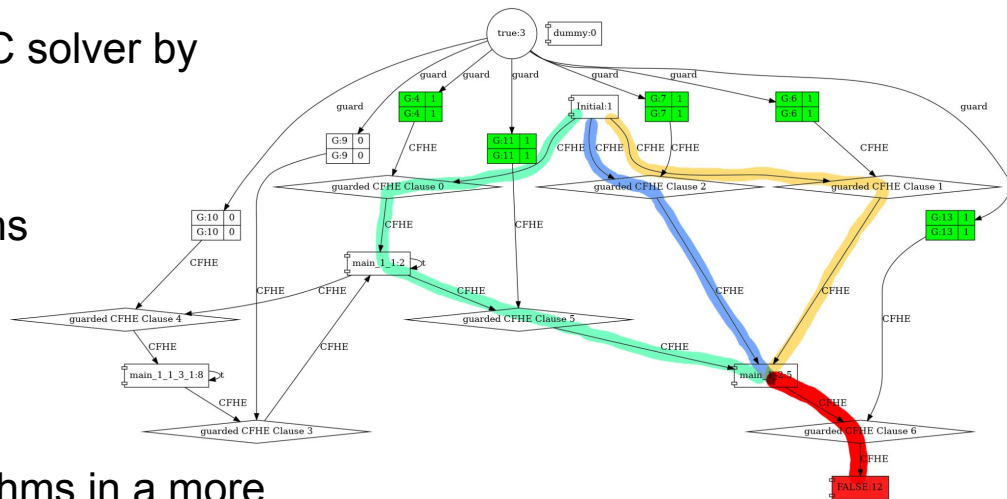
Conclusion

- GNN can be used to speed up CHC solver by predicting MUSes
- GNN learns simple patterns
- It is difficult to learn intricate patterns



Conclusion

- GNN can be used to speed up CHC solver by predicting MUSes
- GNN learns simple patterns
- It is difficult to learn intricate patterns



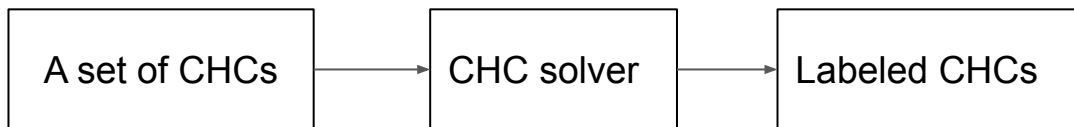
Future work

- Integrating the GNN with the algorithms in a more interactive manner
- Add attention mechanism when train the GNN models

Thank you!

Q & A

Deep Learning-Based Framework (extract training data)



Label

Clauses

1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

When there are multiple MUSes

- Union
- Intersection
- Single

Constraint Horn Clauses (CHCs)

A CHC is a formula in the format

$$\forall V. L[X] \leftarrow L_1[X_1] \wedge \dots \wedge L_n[X_n] \wedge \varphi$$

Where

V are variables,

X_i are terms over V ,

$L, L_1, \dots, L_n$ are n -ary relation symbols,

$L_i[X_i]$ is an atom of relation symbol to the terms,

φ is a constraint in the background theory T .

A set of CHCs (example)

A CHC is a formula in the format

$$\forall V. L[X] \leftarrow L_1[X_1] \wedge \dots \wedge L_n[X_n] \wedge \varphi$$

$$L_1(x) \leftarrow \text{true}$$

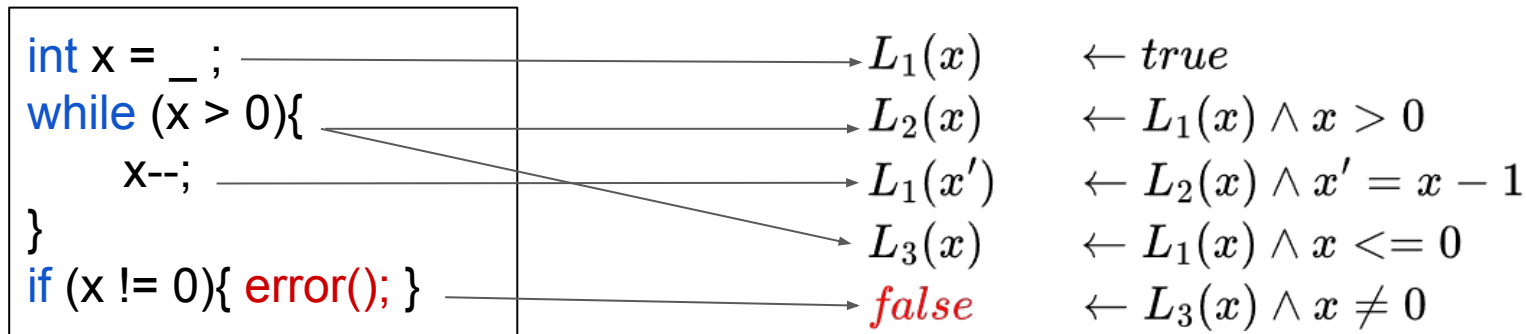
$$L_2(x) \leftarrow L_1(x) \wedge x > 0$$

$$L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$$

$$L_3(x) \leftarrow L_1(x) \wedge x \leq 0$$

$$\text{false} \leftarrow L_3(x) \wedge x \neq 0$$

A program and its Constraint Horn Clauses (CHCs)



Minimal Unsatisfiable Subsets (MUSes) of CHCs

- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$

MUSes of CHCs

- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$

{[1], [4], [5]} is the only MUSes

MUSes of CHCs

- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$

$\{[1], [4], [5]\}$ is the only MUSes

- Algorithms
 - Counterexample-guided abstraction refinement (CEGAR)
 - Symbolic execution (Symex)

MUSes of CHCs

Score

0.8	[1]	$L_1(x)$	$\leftarrow true$
0.2	[2]	$L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3]	$L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4]	$L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5]	$false$	$\leftarrow L_3(x) \wedge x \neq 0$

{[1], [4], [5]} is the only MUSes

- Algorithms
 - Counterexample-guided abstraction refinement (CEGAR)
 - Symbolic execution (Symex)

MUSHyperNet Framework

- Working pipeline
 - Extract train data
 - Represent CHCs by graphs
 - Train Graph Neural Network (GNN) models
 - Guide the algorithms by predicted MUSes of CHCs

Extract train data

- Binary classification label

- Union
- Intersection
- Single

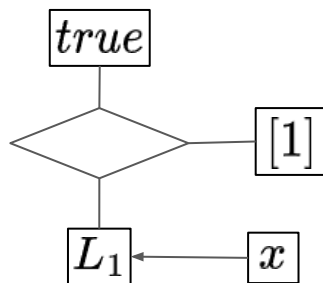
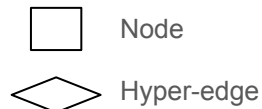
Label	Clauses	
1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

Represent CHCs by graphs

Label

Clauses

1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

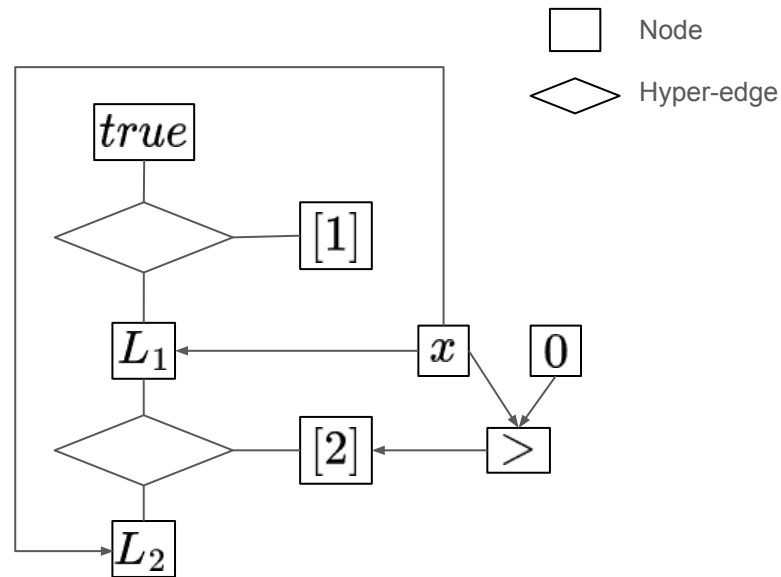


Represent CHCs by graphs

Label

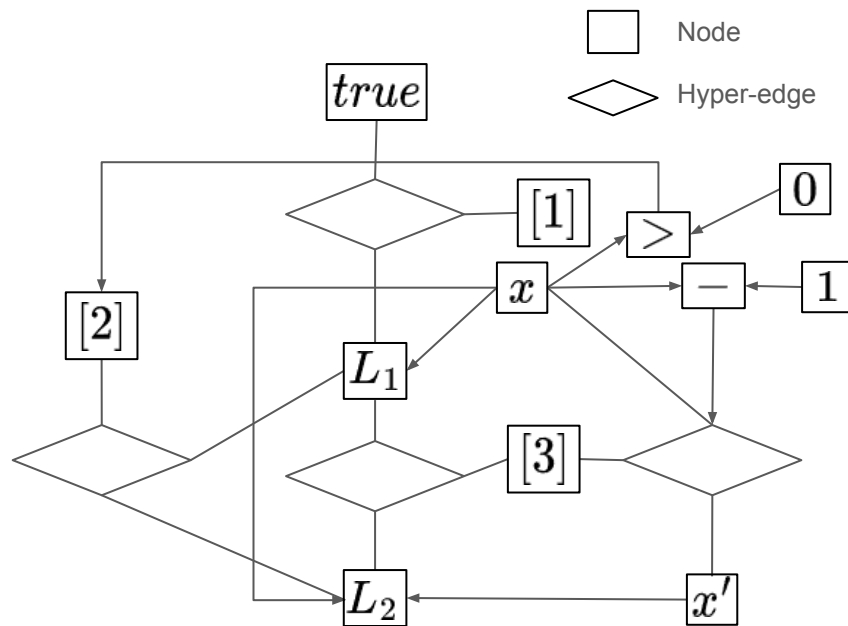
Clauses

1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$



Represent CHCs by graphs

Label	Clauses	
1	[1] $L_1(x)$	$\leftarrow true$
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$

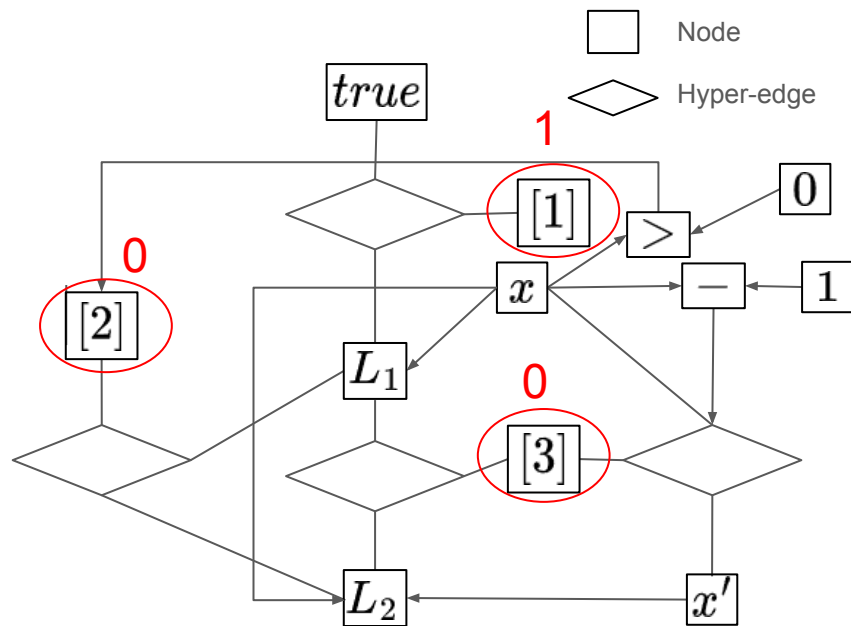


Represent CHCs by graphs

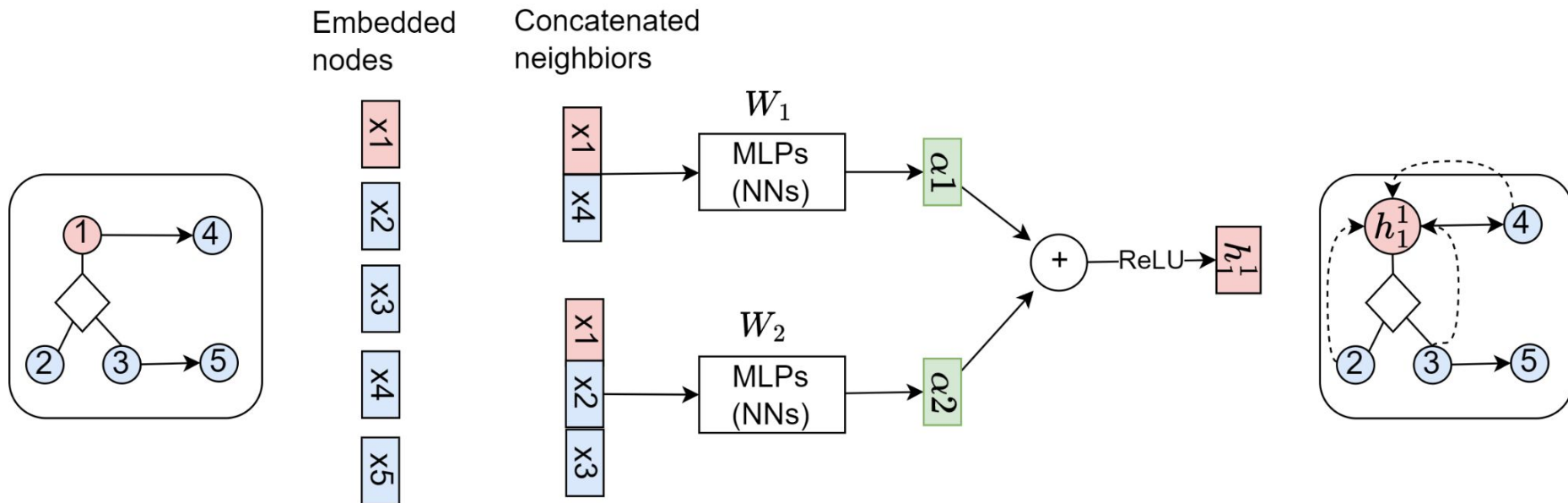
Label

Clauses

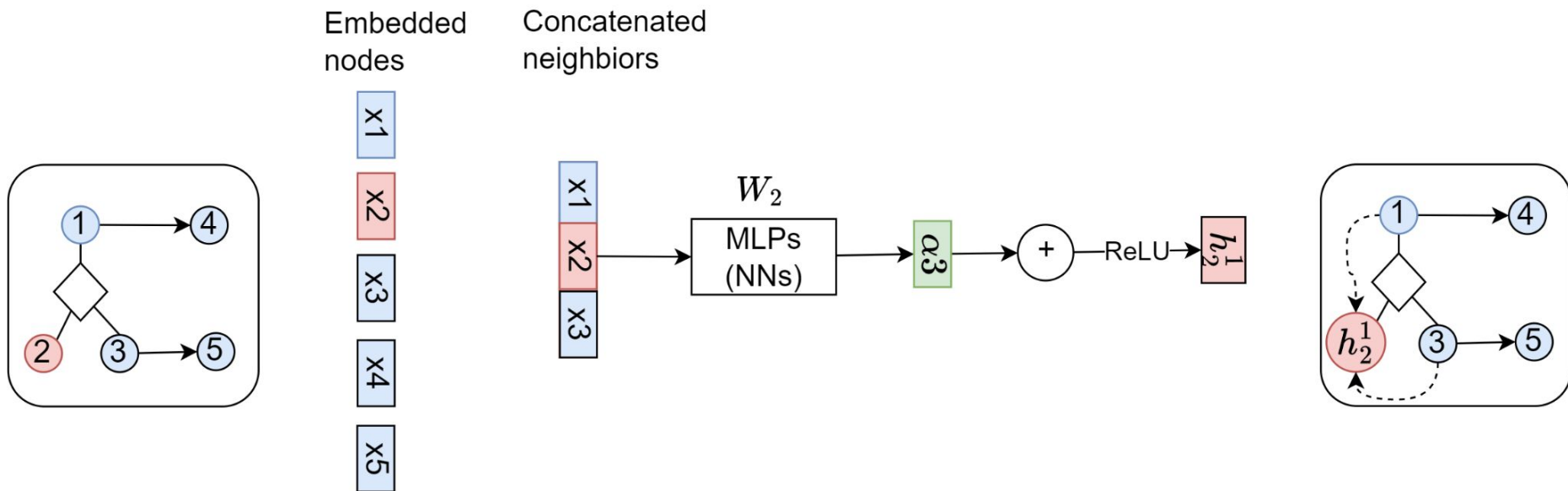
1	[1]	$L_1(x)$	$\leftarrow true$
0	[2]	$L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0	[3]	$L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
1	[4]	$L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
1	[5]	$false$	$\leftarrow L_3(x) \wedge x \neq 0$



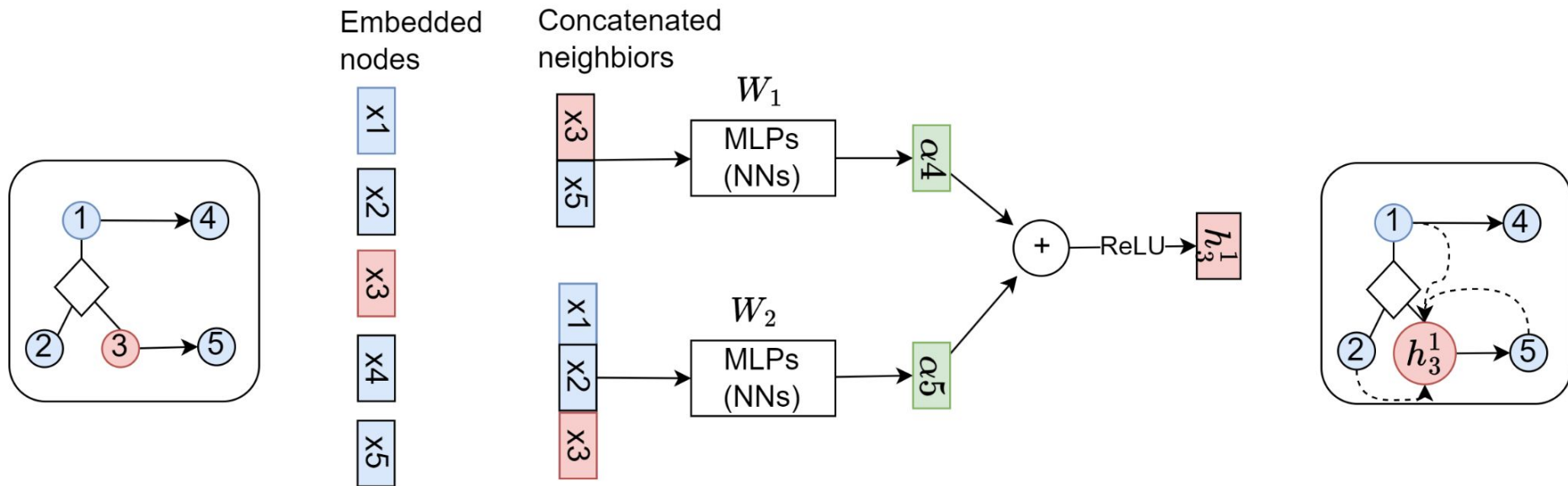
MUSHyperNet Framework (GNN model):



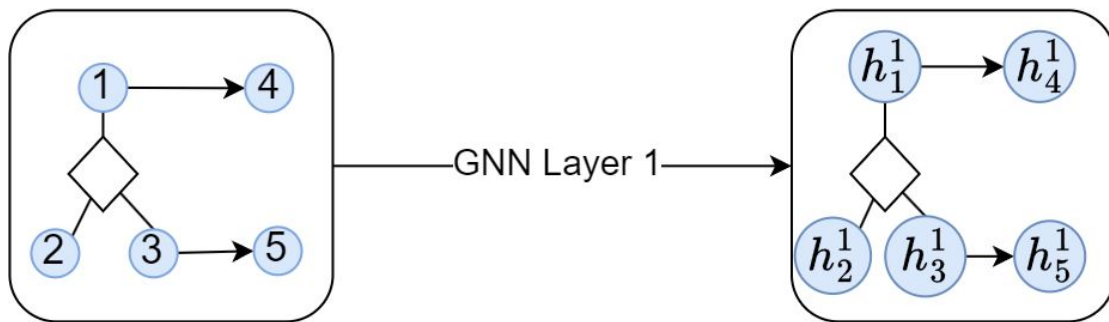
MUSHyperNet Framework (GNN model):



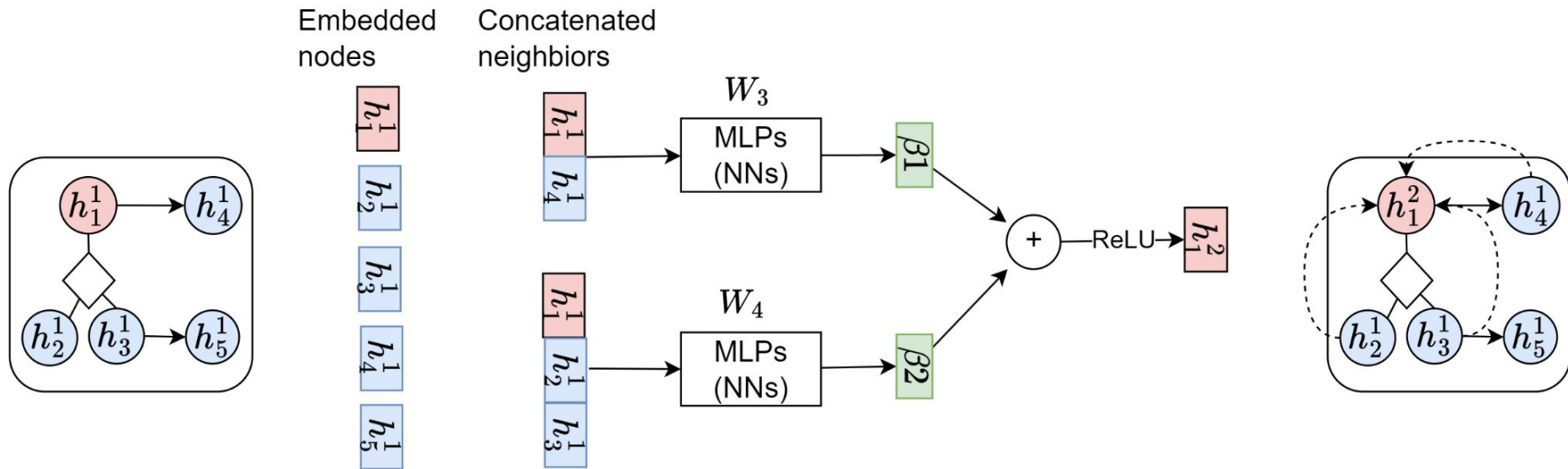
MUSHyperNet Framework (GNN model):



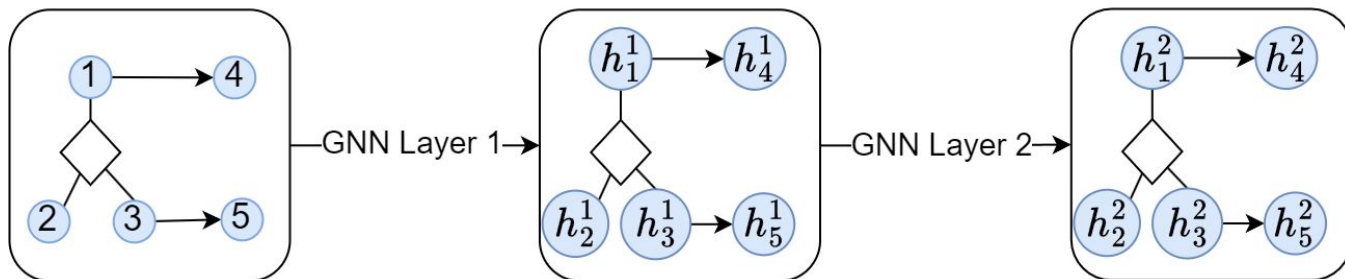
MUSHyperNet Framework (GNN model):



MUSHyperNet Framework (GNN model):



MUSHyperNet Framework (GNN model):





Use predicted MUSes to guide the algorithms

- Prioritize CHCs by using predicted scores of CHCs
 - Use scores alone
 - Combine with original prioritizing scores
 - Add/subtract normalized or ranked scores with coefficient
 - Randomly shift to MUS and original score

Algorithm	Name
CEGAR	Fixed
	Random
	Score
	Rank
	R-Plus
	S-Plus
	R-Minus
SymEx	S-Minus
	Fixed
	Random
	Score
	Rank
	R-Plus
	S-Plus
	R-Minus
	S-Minus
	Two-queue

Experimental results

- Benchmarks from CHC-COMP

Linear LIA problems					Non-linear LIA problems				
8705					8425				
Benchmarks for training			Holdout set		Benchmarks for training			Holdout set	
7834 (90%)			871 (10%)		7579 (90%)			846 (10%)	
UNSAT	SAT	T/O	Eval.	N/A	UNSAT	SAT	T/O	Eval.	N/A
1585	4004	2245	383	488	3315	4010	254	488	358
Train	Valid	N/A			Train	Valid	N/A		
782	87	716			1617	180	1518		

Experimental results (Improved percentage)

Benchmark Algorithm	MUS data set (best count)	Best ranking function (improvement in %)						
		Number of Solved Problems			Average Time			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Linear CEGAR	Union	R-Plus	R-Plus	R-Minus	R-Plus	S-Plus	S-Minus	Rank
	(0)	(1.4%)	(2.4%)	(1.0%)	(1.3%)	(19.1%)	(46.5%)	(31.1%)
	Single	Rank	R-Plus	Rank	R-Plus	S-Plus	R-Minus	Rank
	(3)	(3.6%)	(4.0%)	(8.2%)	(1.9%)	(26.6%)	(57.9%)	(36.3%)
	Intersection	R-Plus	S-Plus	R-Plus	R-Plus	S-Plus	R-Minus	S-Plus
	(4)	(4.1%)	(0.8%)	(9.3%)	(3.1%)	(27.6%)	(45.0%)	(0.0%)
Linear SymEx	Union	Two-Q	S-Plus*	Random	Two-Q	R-Minus	R-Minus	S-Plus
	(4)	(1.0%)	(0.0%)	(2.0%)	(0.9%)	(12.7%)	(30.2%)	(26.5%)
	Single	S-Minus*	S-Plus*	Random	Random	S-Plus	Random	S-Plus
	(3)	(0.5%)	(0.0%)	(2.0%)	(0.8%)	(12.9%)	(28.4%)	(17.6%)
	Intersection	S-Plus*	S-Plus*	S-Plus*	S-Plus	Score	Random	R-Plus
	(5)	(1.0%)	(0.0%)	(2.0%)	(1.3%)	(9.5%)	(28.4%)	(35.8%)

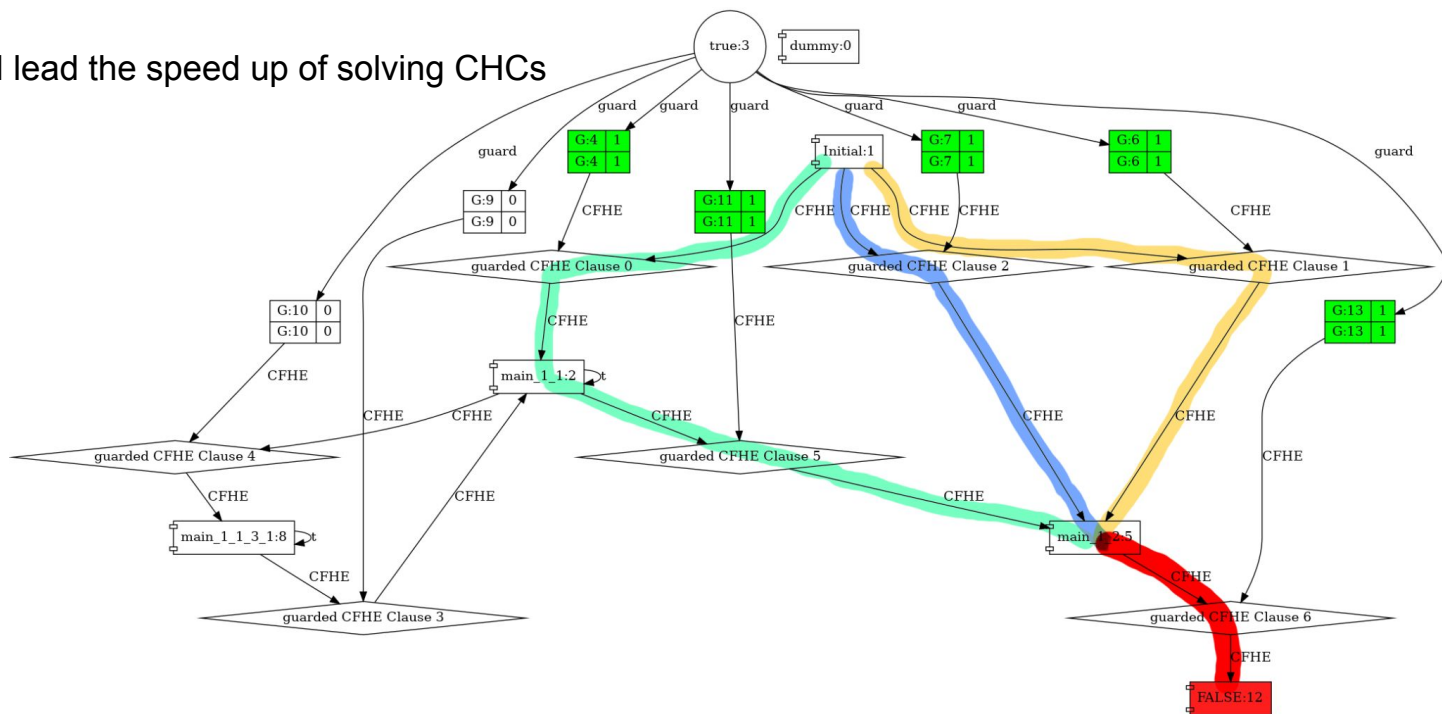
Experimental results (Improved percentage)

Benchmark Algorithm	MUS data set (best count)	Best ranking function (improvement in %)						
		Number of Solved Problems			Average Time			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Non-Linear CEGAR	Union (7)	S-Plus (0.5%)	S-Plus (0.8%)	S-Plus* (0.0%)	S-Plus (7.1%)	R-Minus (20.8%)	Rank (53.5%)	S-Plus (19.4%)
	Single (1)	R-Plus (0.2%)	R-Plus (0.4%)	R-Plus* (0.0%)	R-Plus (6.6%)	S-Plus (18.4%)	R-Minus (52.8%)	R-Minus (14.2%)
	Intersection (1)	R-Plus* (0.0%)	S-Plus (0.5%)	S-Plus* (0.0%)	R-Plus (5.9%)	R-Plus (20.3%)	Rank (45.8%)	S-Plus (16.8%)
Non-Linear SymEx	Union (6)	Two-Q (6.1%)	S-Minus* (1.6%)	Random (12.3%)	Two-Q (13.3%)	R-Minus (7.3%)	Score (5.1%)	R-Plus (19.9%)
	Single (3)	Two-Q (6.1%)	Score (1.6%)	Two-Q (12.9%)	Two-Q (12.4%)	Rank (-2.2%)	R-Minus (0.2%)	Two-Q (11.2%)
	Intersection (3)	Two-Q (6.1%)	S-Plus (1.6%)	Two-Q (12.9%)	Two-Q (12.7%)	S-Minus (0.6%)	Two-Q (1.7%)	S-Plus (5.4%)

Conclusion

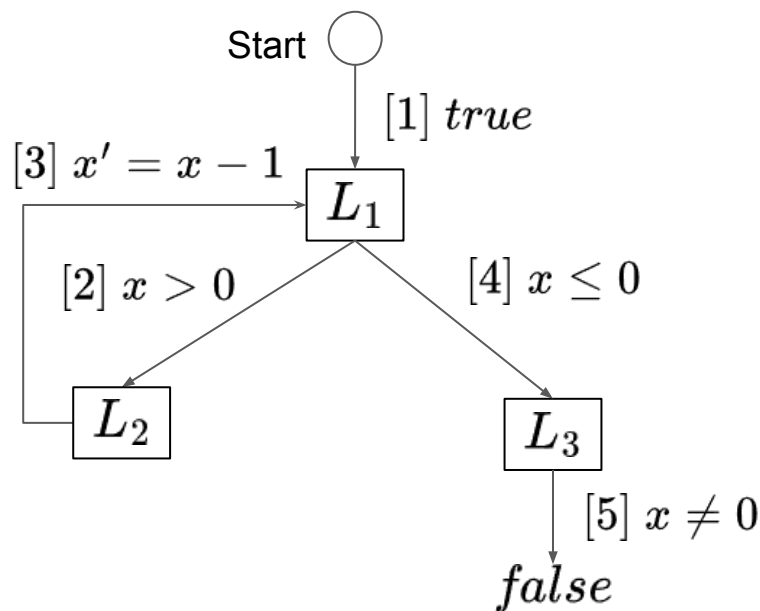
- GNN can be used lead the speed up of solving CHCs

- Future works
-



Visualize CHCs with dependency graph

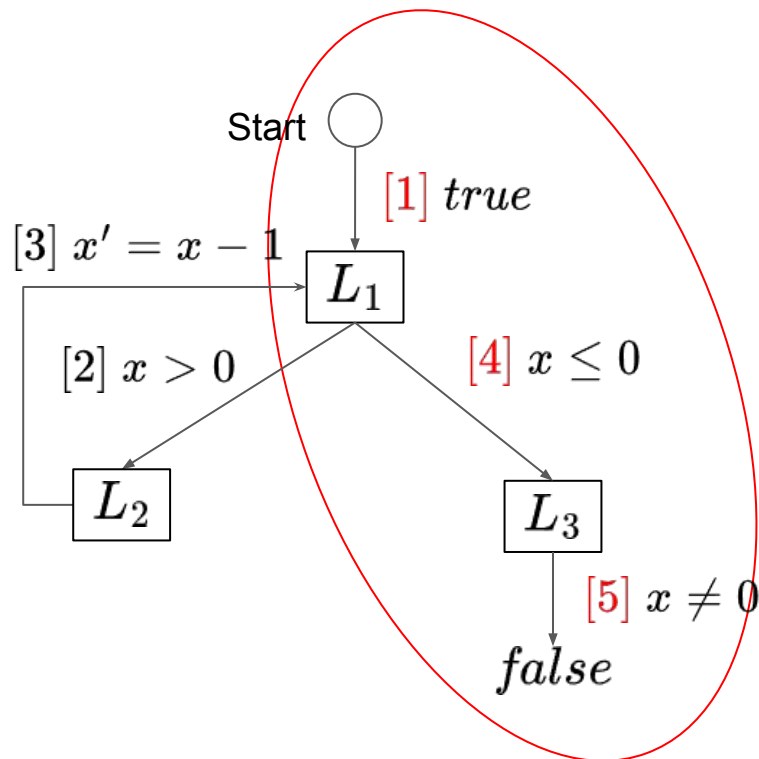
- [1] $L_1(x) \leftarrow true$
- [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
- [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
- [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
- [5] $false \leftarrow L_3(x) \wedge x \neq 0$



MUSes of CHCs

- [1] $L_1(x) \leftarrow true$
 [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
 [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
 [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
 [5] $false \leftarrow L_3(x) \wedge x \neq 0$

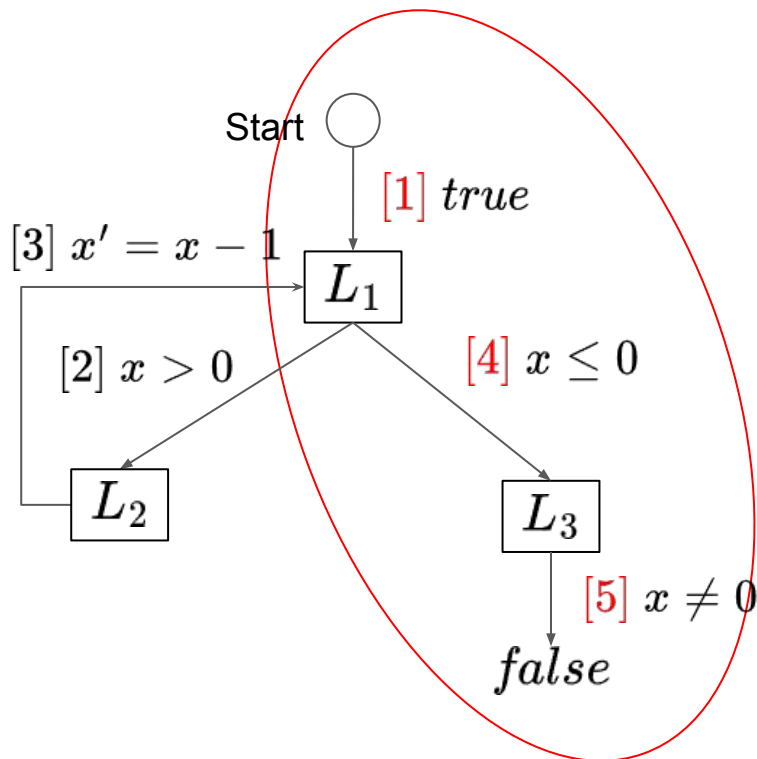
$\{[1], [4], [5]\}$ is the only MUSes



MUSes of CHCs

[1] $L_1(x) \leftarrow true$
 [2] $L_2(x) \leftarrow L_1(x) \wedge x > 0$
 [3] $L_1(x') \leftarrow L_2(x) \wedge x' = x - 1$
 [4] $L_3(x) \leftarrow L_1(x) \wedge x \leq 0$
 [5] $false \leftarrow L_3(x) \wedge x \neq 0$

- Algorithms
 - Counterexample-guided abstraction refinement (CEGAR)
 - Symbolic execution (Symex)

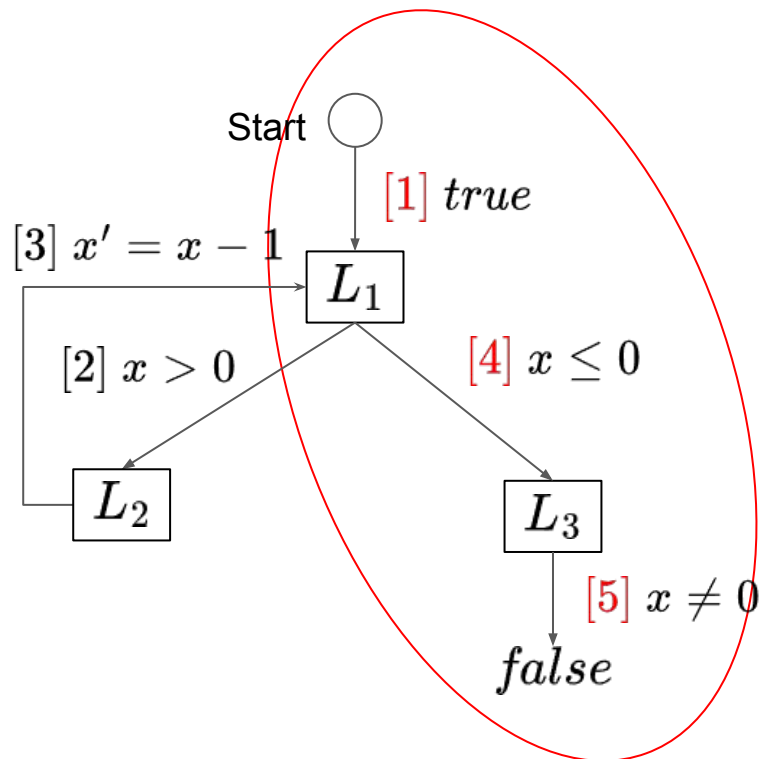


MUSes of CHCs

Score

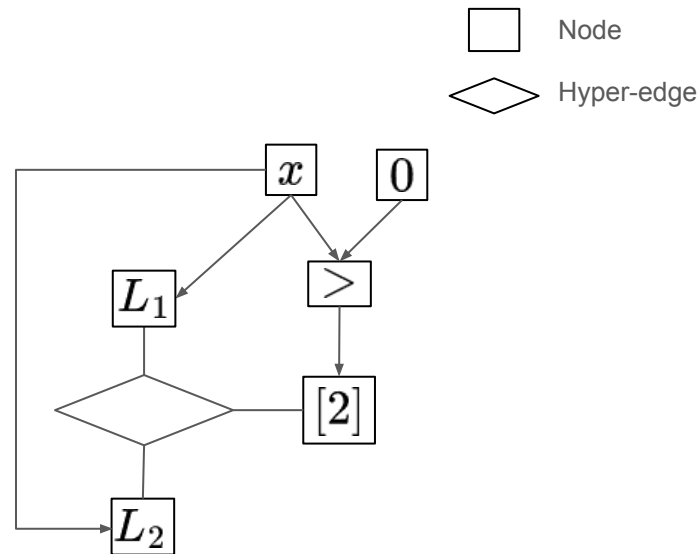
0.8	[1]	$L_1(x)$	$\leftarrow true$
0.2	[2]	$L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$
0.1	[3]	$L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$
0.75	[4]	$L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$
0.6	[5]	$false$	$\leftarrow L_3(x) \wedge x \neq 0$

- Algorithms
 - Counterexample-guided abstraction refinement (CEGAR)
 - Symbolic execution (Symex)



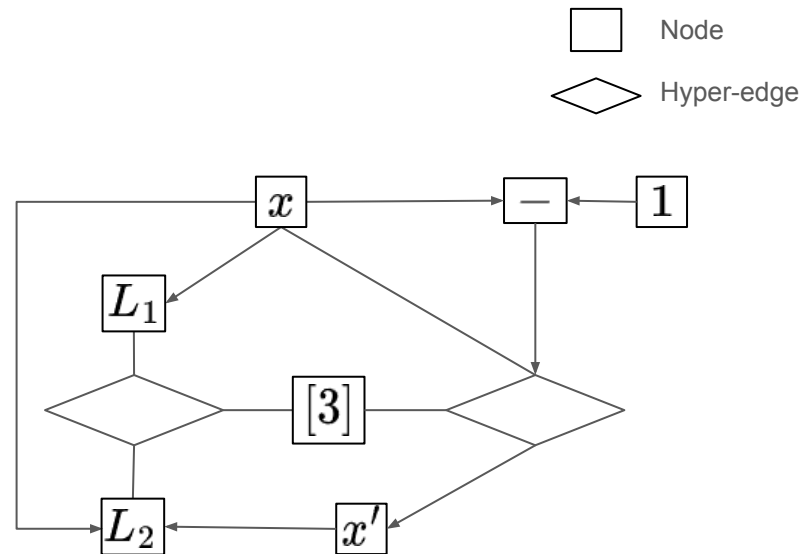
Represent CHCs by graphs

Label	Clauses		
1	[1] $L_1(x)$	$\leftarrow true$	
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$	
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$	
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$	
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$	



Represent CHCs by graphs

Label	Clauses		
1	[1] $L_1(x)$	$\leftarrow true$	
0	[2] $L_2(x)$	$\leftarrow L_1(x) \wedge x > 0$	
0	[3] $L_1(x')$	$\leftarrow L_2(x) \wedge x' = x - 1$	
1	[4] $L_3(x)$	$\leftarrow L_1(x) \wedge x \leq 0$	
1	[5] $false$	$\leftarrow L_3(x) \wedge x \neq 0$	



Experimental results

Benchmark Algorithm	Ranking Function	Number of Solved Problems (improvement %)			Average Time (improvement %)			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Non Linear CEGAR	Default	432	250	182	131.12	42.05	43.34	40.28
	Random	425	243	182	143.42	34.27	34.84	38.75
		(-1.6%)	(-2.8%)	(0.0%)	(-9.4%)	(-11.1%)	(19.6%)	(3.8%)
	R-Plus	432	250	182	122.29	31.74	28.59	37.82
		(0.0%)	(0.0%)	(0.0%)	(6.7%)	(17.8%)	(34.0%)	(6.1%)
	R-Minus	417	240	177	154.07	26.20	21.46	32.51
		(-3.5%)	(-4.0%)	(-2.7%)	(-17.5%)	(20.8%)	(50.5%)	(19.3%)
	S-Plus	434	252	182	121.75	34.64	35.97	39.10
		(0.5%)	(0.8%)	(0.0%)	(7.1%)	(13.1%)	(17.0%)	(2.9%)
	S-Minus	421	242	179	149.02	31.76	26.33	38.95
Non Linear SymEx	Portfolio	435	253	182	113.49	28.24	30.57	31.75
		(0.7%)	(1.2%)	(0.0%)	(13.4%)	(29.1%)	(29.5%)	(21.2%)
	Default	342	187	155	343.82	28.39	29.05	27.59
	Random	362	188	174	301.90	32.67	36.24	41.83
		(5.8%)	(0.5%)	(12.3%)	(12.2%)	(-15.1%)	(-24.8%)	(-51.6%)
	R-Plus	339	190	149	357.18	27.88	47.71	22.10
		(-0.9%)	(1.6%)	(-3.9%)	(-3.9%)	(0.3%)	(-64.2%)	(19.9%)
	R-Minus	361	189	172	299.86	26.35	37.68	27.98
		(5.6%)	(1.1%)	(11.0%)	(12.8%)	(7.3%)	(-29.7%)	(-1.4%)
	S-Plus	340	189	151	352.84	29.04	41.41	24.54
		(-0.6%)	(1.1%)	(-2.6%)	(-2.6%)	(-0.3%)	(-42.5%)	(11.1%)
Non Linear SymEx	S-Minus	362	190	172	303.65	28.62	44.11	37.95
		(5.8%)	(1.6%)	(11.0%)	(11.7%)	(-0.4%)	(-51.8%)	(-37.5%)
	Two-queue	363	189	174	297.93	30.15	41.14	32.51
Non Linear SymEx	Portfolio	(6.1%)	(1.1%)	(12.3%)	(13.3%)	(-6.2%)	(-41.6%)	(-17.8%)
		366	191	175	288.85	22.29	42.42	26.75
	Portfolio	(7.0%)	(2.1%)	(12.9%)	(16.0%)	(21.4%)	(-46.0%)	(3.0%)

Experimental results

- At least one setting has improvement

Benchmark Algorithm	Ranking Function	Number of Solved Problems (improvement %)			Average Time (improvement %)			
		Total	SAT	UNSAT	All	Common	SAT	UNSAT
Linear CEGAR	Default	222	125	97	519.38	25.77	38.97	8.77
	Random	221 (-0.5%)	124 (-0.8%)	97 (0.0%)	523.58 (-0.8%)	27.49 (-29.5%)	37.05 (4.9%)	15.85 (-80.7%)
	R-Plus	225 (1.4%)	128 (2.4%)	97 (0.0%)	512.41 (1.3%)	21.65 (16.0%)	42.89 (-10.1%)	11.99 (-36.7%)
	R-Minus	220 (-0.9%)	122 (-2.4%)	98 (1.0%)	526.08 (-1.3%)	18.02 (-24.4%)	30.93 (20.6%)	21.60 (-146.3%)
	S-Plus	222 (0.0%)	125 (0.0%)	97 (0.0%)	517.43 (0.4%)	20.92 (19.1%)	34.13 (12.4%)	7.32 (16.5%)
	S-Minus	219 (-1.4%)	122 (-2.4%)	97 (0.0%)	522.97 (-0.7%)	12.56 (2.4%)	20.86 (46.5%)	9.81 (-11.9%)
	Portfolio	229 (3.2%)	130 (4.0%)	99 (2.1%)	503.16 (3.1%)	18.28 (29.1%)	45.67 (-17.2%)	19.94 (-127.4%)
	Default	200	101	99	590.68	33.16	55.42	10.44
	Random	201 (0.5%)	100 (-1.0%)	101 (2.0%)	586.12 (0.8%)	30.08 (-8.5%)	39.69 (28.4%)	20.95 (-100.7%)
	R-Plus	192 (-4.0%)	101 (0.0%)	91 (-8.1%)	617.60 (-4.6%)	38.59 (-10.9%)	52.87 (4.6%)	21.99 (-110.6%)
Linear SymEx	R-Minus	200 (0.0%)	100 (-1.0%)	100 (1.0%)	586.24 (0.8%)	24.67 (12.7%)	38.69 (30.2%)	10.60 (-1.5%)
	S-Plus	198 (-1.0%)	101 (0.0%)	97 (-2.0%)	595.02 (-0.7%)	30.22 (11.6%)	50.97 (8.0%)	7.67 (26.5%)
	S-Minus	201 (0.5%)	101 (0.0%)	100 (1.0%)	586.35 (0.7%)	30.64 (7.8%)	50.57 (8.8%)	10.65 (-2.0%)
	Two-queue	202 (1.0%)	101 (0.0%)	101 (2.0%)	585.58 (0.9%)	35.11 (-5.9%)	49.94 (9.9%)	20.14 (-92.9%)
	Portfolio	206 (3%)	101 (0.0%)	105 (6.1%)	569.1 (3.7%)	25.79 (22.2%)	44.58 (19.6%)	10.16 (2.6%)
	Default	200	101	99	590.68	33.16	55.42	10.44
	Random	201 (0.5%)	100 (-1.0%)	101 (2.0%)	586.12 (0.8%)	30.08 (-8.5%)	39.69 (28.4%)	20.95 (-100.7%)
	R-Plus	192 (-4.0%)	101 (0.0%)	91 (-8.1%)	617.60 (-4.6%)	38.59 (-10.9%)	52.87 (4.6%)	21.99 (-110.6%)

Minimal Unsatisfiable Subsets (MUSes) of CHCs

