NEURAL INFORMATION PROCESSING SYSTEMS
38th Conference on Neural Information Processing Systems (NeurIPS 2024)

Ordering-Based Causal Discovery for Linear and Nonlinear Relations

(Presenter)
Zhuopeng Xu, Yujie Li, Cheng Liu, Ning Gui*

中南大學
CENTRAL SOUTH UNIVERSITY

Background - Causal Discovery with Observational Data

Causal discovery uncovers latent causal relationships within data by modeling a Directed Acyclic Graph (DAG) connecting various variables.

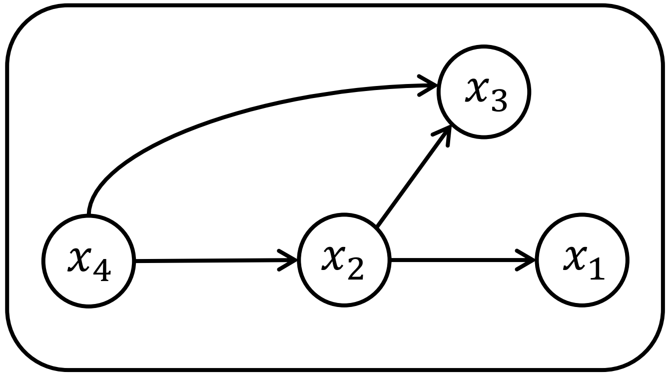
Input: data with n samples and d features

x_1	…	x_2	x_3	…	x_4
0.5	…	1.0	3.0	…	1.0
3.5	…	4.0	8.0	…	2.0
1.75	…	2.25	5.08	…	1.5
…	…	…	…	…	…
-0.5	…	0.0	1.0	…	0.0

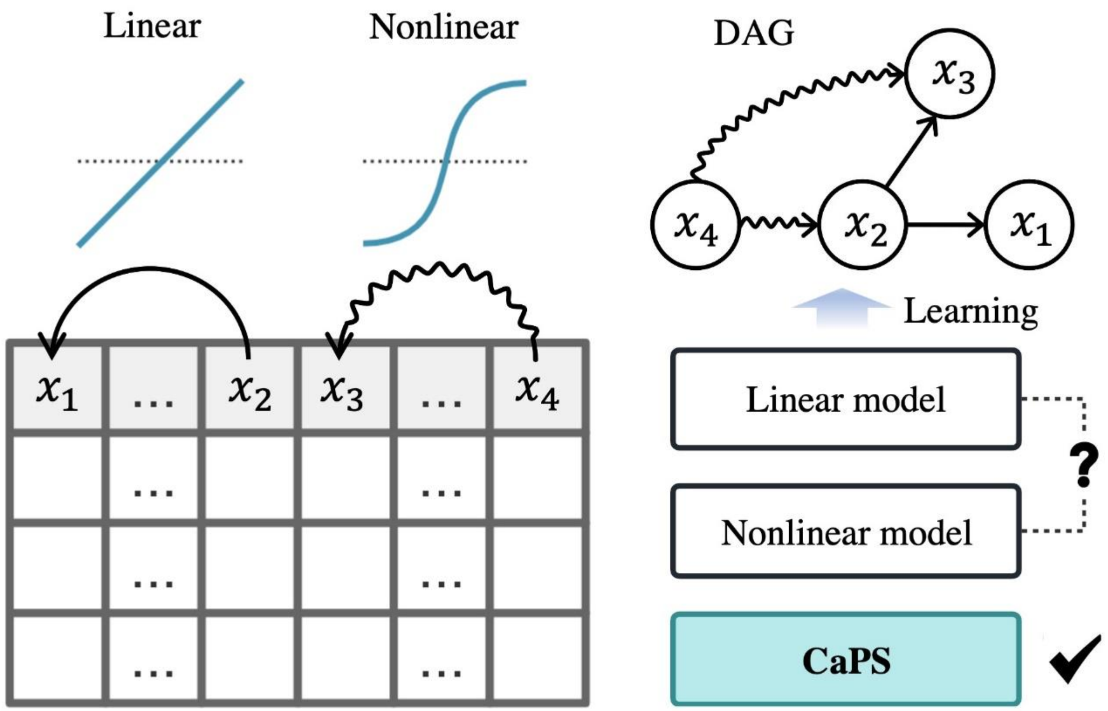
By observation only 👀

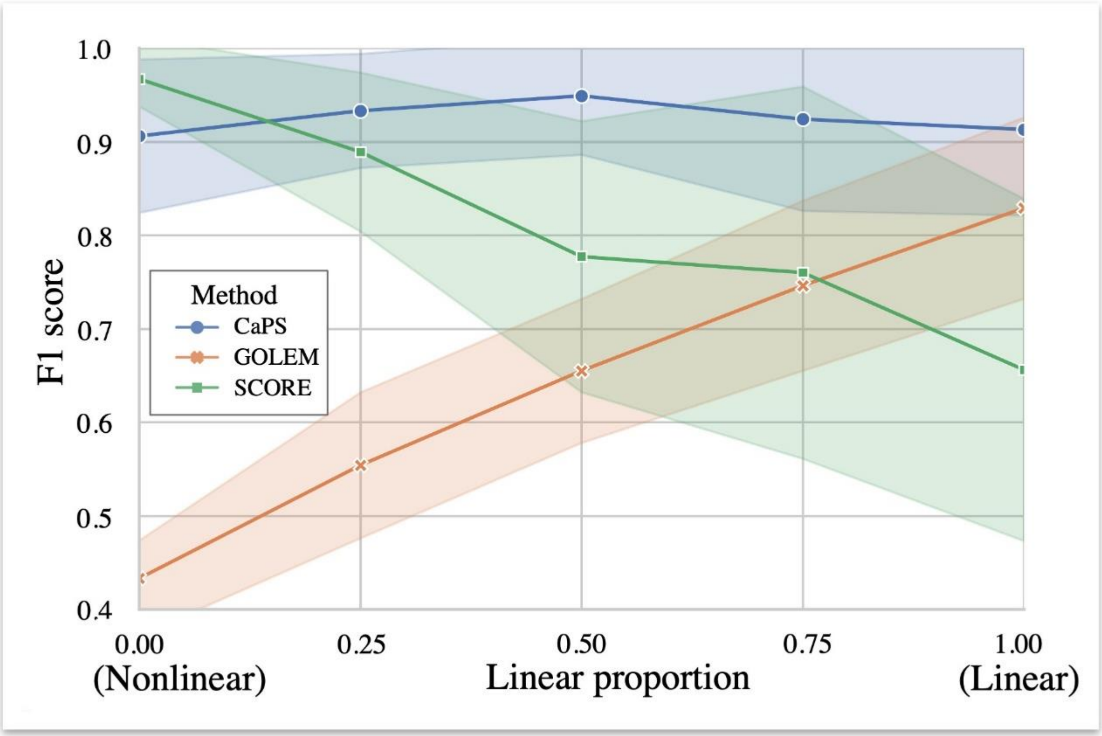
Without any intervention 🫷

Output: DAG with d nodes



Motivation - Linear and Nonlinear Challenge





- Existing approaches normally limit their discussions to pure **linear or nonlinear** relations, which will suffer significant performance loss when their assumptions mismatch.
- Since we don't know whether the real-world data is linear or nonlinear, it is difficult to choose an effective model. Thus, we need a method that works well in both **linear** and **nonlinear** and most possibly **mixed** cases.

Preliminaries - Ordering-based Causal Discovery

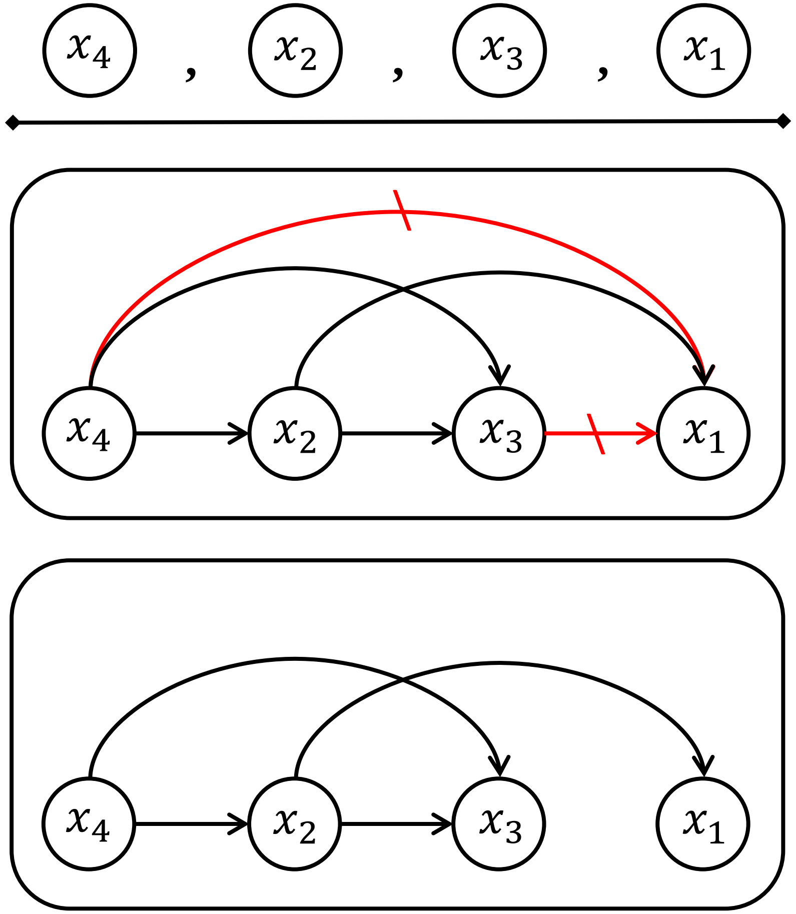
Input:
data with n samples and d features

x_1	…	x_2	x_3	…	x_4
0.5	…	1.0	3.0	…	1.0
3.5	…	4.0	8.0	…	2.0
1.75	…	2.25	5.08	…	1.5
…	…	…	…	…	…
-0.5	…	0.0	1.0	…	0.0

Stage 1:
Estimate topological ordering

Stage 2:
Initialize DAG & Pruning

Output:
DAG with d nodes



This two-stage strategy has been shown to have the capability to reduce the complexity of DAG discovery while keeping the acyclic constraint.

Methods – Topological Ordering (Stage 1 of CaPS)

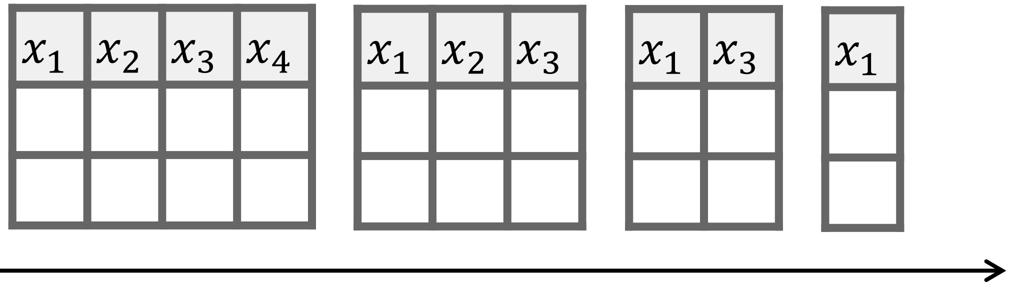
A novel ordering criterion for leaf nodes discrimination 🍃

Theorem 1. *Let* $s(x) = \nabla \log p(x)$ *be the score and let* $diag(\cdot)$ *be the diagonal elements of the matrix. For any* x_j *in the causal graph* $\mathcal{G}$*:*

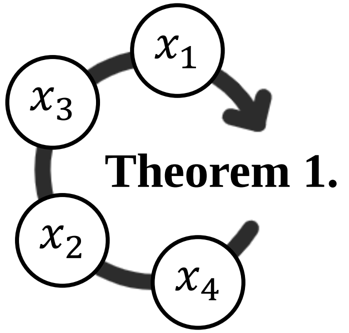
$$j = \operatorname{argmax}\left(diag\left(\mathbb{E}\left[\frac{\partial s(x)}{\partial x}\right]\right)\right) \Rightarrow x_j \text{ is a leaf node}$$

Proof. *See section 4.1 and appendix A.2 in our manuscript.*

x_1	…	x_2	x_3	…	x_4
0.5	…	1.0	3.0	…	1.0
3.5	…	4.0	8.0	…	2.0
1.75	…	2.25	5.08	…	1.5
…	…	…	…	…	…
-0.5	…	0.0	1.0	…	0.0



Iteratively remove current leaf node and estimate the expectation of the diagonal of the score's Jacobian



Methods – Topological Ordering (Stage 1 of CaPS)

Two identifiable scenarios

In this paper, we give two sufficient conditions for causal identifiability **without any assumption of causal relations, i.e., linear or nonlinear assumption.**

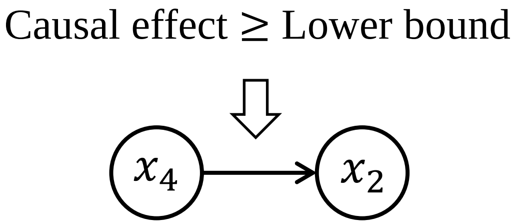
(i) Non-decreasing variance of noises.
For any two noises ϵ_i and ϵ_j, $\sigma_j \geq \sigma_i$, if $\pi(i) < \pi(j)$.

noises: $\sigma_4 \leq \sigma_2 \leq \sigma_3 \leq \sigma_1$

features: x_4 , x_2 , x_3 , x_1

(ii) Non-weak causal effect.

For any non-leaf nodes x_j, $\sum_{i \in Ch(j)} \frac{1}{\sigma^2} \mathbb{E}\left[\left(\frac{\partial f_i}{\partial x_j}(pa_i(x))\right)^2\right] \geq \frac{1}{\sigma_{min}^2} - \frac{1}{\sigma_j^2}$



Conditions *(i)* and *(ii)* are two different identifiable scenarios, and CaPS only needs **one of them** to be satisfied. (see Assumption 1 for more details)

Methods – Parent Score (Stage 2 of CaPS)

CENTRAL SOUTH UNIVERSITY 中南大學　NEURAL INFORMATION PROCESSING SYSTEMS

A new metric to approximate the average causal effect

- *Can we utilize some of the information hidden in Theorem 1 for further post-processing?*

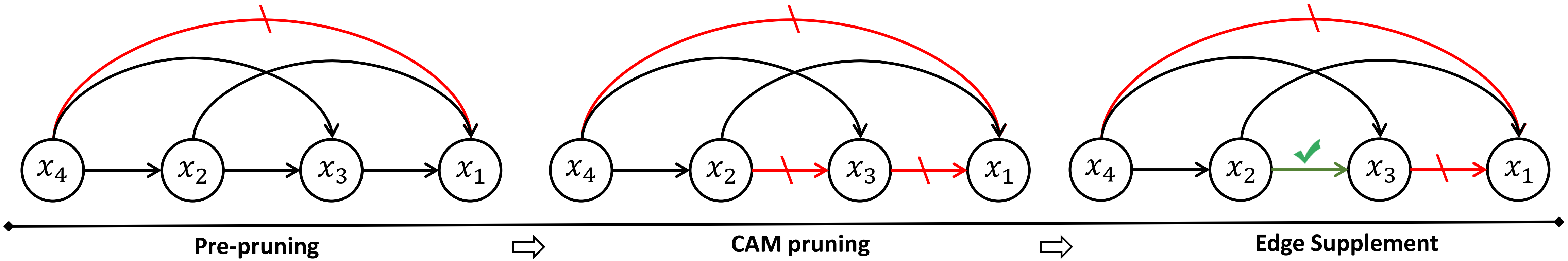
Parent score is introduced to **reflects the strength of the average causal effect** of a given parent. This metric can be obtained directly by decoupling from Theorem 1 **without any additional computational complexity.**

Theorem 1. ⇒ **Parent Score.**

$$\mathcal{P}_{i,j} = \begin{cases} \frac{1}{\sigma_i^2}\mathbb{E}\left[\left(\frac{\partial f_i}{\partial x_j}(pa_i(x))\right)^2\right], & x_j \in pa_i(x) \\ 0, & x_j \notin pa\_i\ (x) \end{cases}$$

Pre-pruning. Remove the low-confidence edges and reduce the searching space.

Edge Supplement. Use high-confidence parents to supplement the edge.



Experiments – Different linear proportion & Order divergence

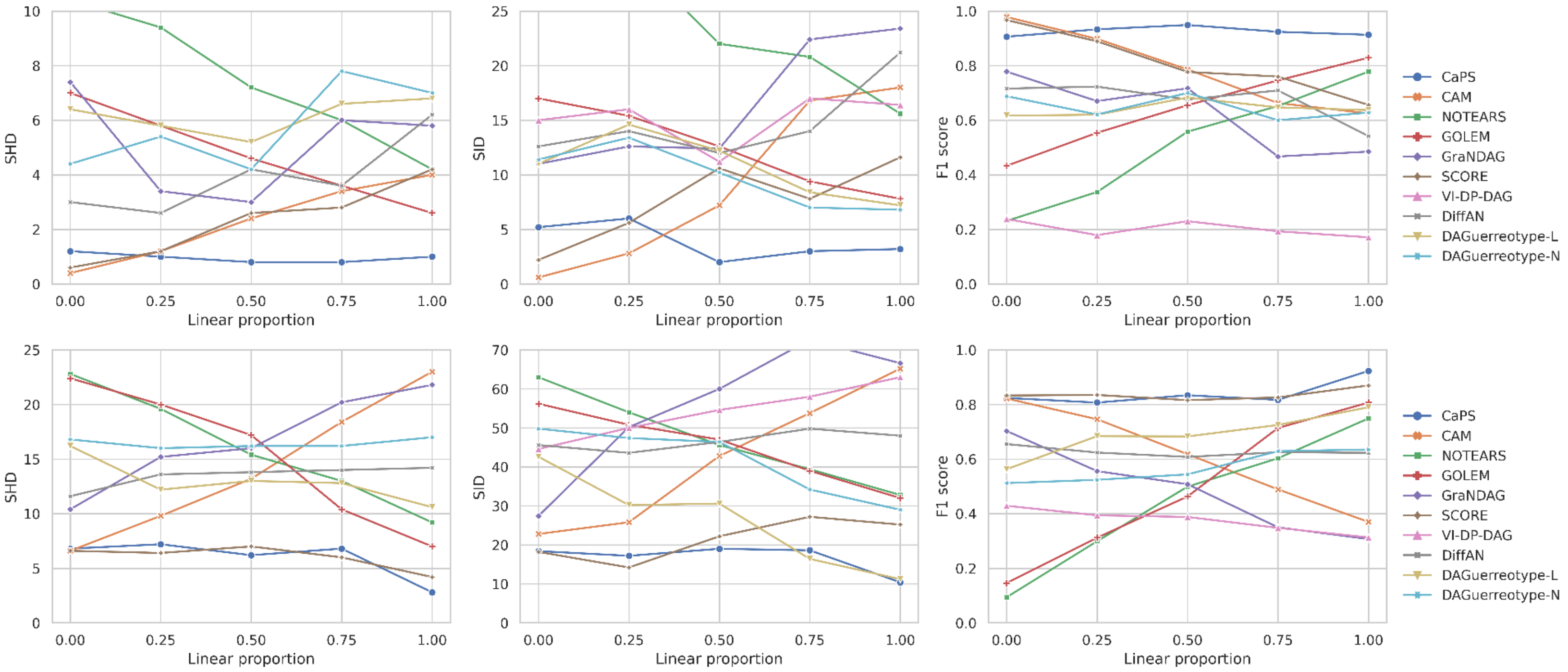


Figure 1: Results of SynER1 (top) and SynER4 (bottom) with different linear proportions, where linear proportion equal to 0.0 means all relations are nonlinear and 1.0 means all relations are linear.

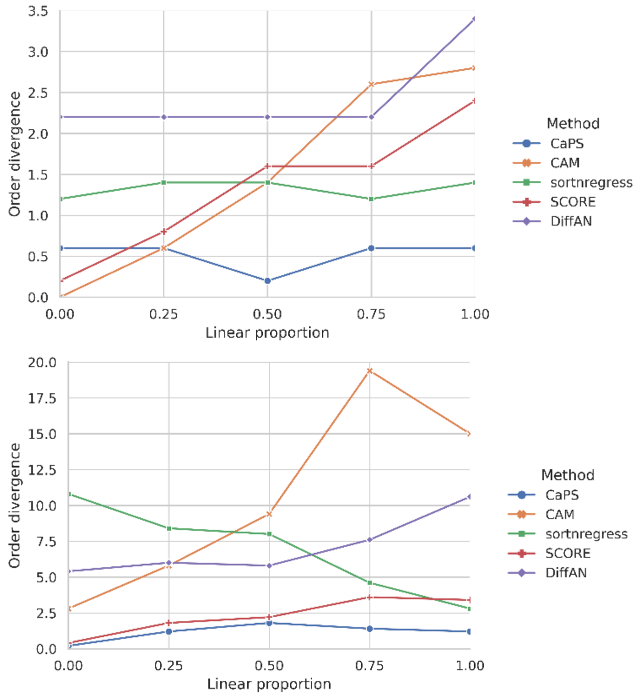


Figure 2: Order Divergence of SynER1 (top) and SynER4 (bottom) with different linear proportions and sparsity.

- CaPS performs better for both sparser (SynER1) and denser (SynER4) graphs under almost **all linear proportion** in SHD, SID and F1.
- Compared to other ordering-based method, CaPS consistently has the best or a competitive **order divergence**.

Experiments – Larger-scale datasets & actual-time cost

CaPS consistently achieves best performance in **larger-scale** causal graph while its **time cost is competitive**.

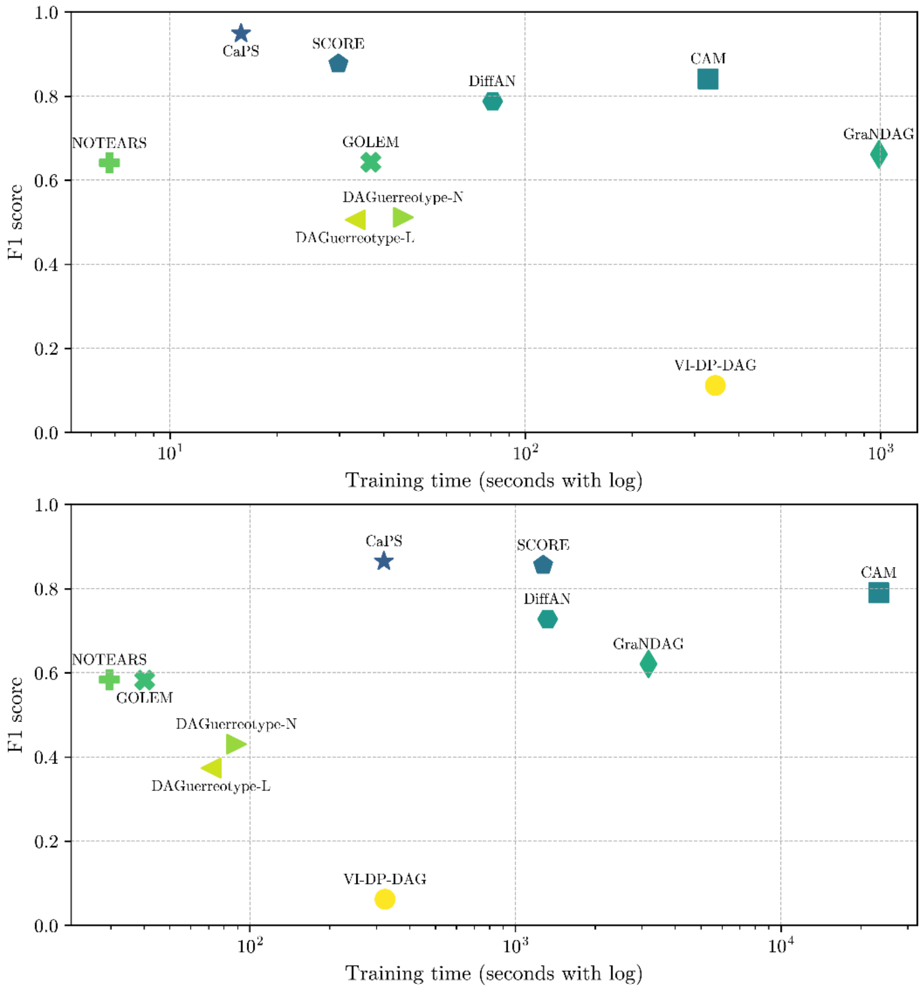


Figure 3: F1 score and training time of SynER1 with larger-scale causal graph.

Prop.	Metrics	NOTEARS	COLEM	GraNDAG	CAM	VI-DP-DAG	SCORE	DiffAN	DAGu.-L	DAGu.-N	CaPS
SynER1 (d=20)											
	SHD	17.4±1.7	15.2±1.8	5.3±1.2	1.4±0.8	123.2±4.9	1.4±0.8	6.2±3.7	23.6±2.0	18.3±2.6	**0.8±0.4**
0	SID	81.2±16.1	81.0±22.6	18.6±7.7	8.4±6.1	33.4±7.3	5.4±3.2	31.6±17.5	63.3±23.5	46.6±22.8	**3.4±2.8**
	F1	27.0±11.1	41.3±8.9	82.0±6.3	95.5±2.6	18.4±2.1	96.5±2.3	79.9±11.4	44.0±10.8	56.8±6.6	**98.1±1.0**
	SHD	14.8±1.6	14.2±2.0	5.3±0.9	3.4±1.3	118.4±7.4	**0.4±0.4**	4.6±2.0	18.0±1.6	19.3±3.2	1.2±1.1
0.25	SID	73.8±14.3	78.0±21.7	32.0±9.8	18.2±10.0	56.8±23.9	**3.8±7.6**	32.2±15.0	39.0±2.1	64.0±33.9	7.6±10.2
	F1	44.1±5.7	43.9±12.7	79.8±6.0	87.1±5.4	16.4±3.1	**98.6±1.8**	81.9±8.5	61.5±3.7	52.6±10.2	96.0±3.9
	SHD	10.8±1.7	10.0±1.6	9.6±2.6	5.4±2.6	125.4±8.3	4.6±3.9	6.6±2.6	25.3±6.6	27.0±5.0	**4.2±3.0**
0.5	SID	56.6±13.3	53.6±16.9	63.3±21.6	21.0±9.5	85.2±20.0	17.2±13.3	24.2±12.5	72.3±37.0	40.6±9.7	**17.0±11.4**
	F1	64.1±7.5	64.3±8.4	66.2±6.1	81.4±6.7	11.2±2.2	87.7±9.9	78.8±09.1	50.6±7.6	51.2±0.2	**94.9±4.6**
	SHD	8.8±1.7	7.8±2.9	16.6±3.3	10.2±2.6	119.6±7.3	3.8±2.3	9.0±3.3	21.6±10.2	26.3±1.2	**1.6±1.6**
0.75	SID	46.6±10.4	45.6±21.6	67.6±10.4	57.2±18.8	83.4±44.2	17.4±13.6	41.6±13.3	49.3±36.5	56.6±3.3	**11.8±10.9**
	F1	72.5±4.3	7.14±11.0	43.4±13.6	63.1±10.7	13.8±4.9	88.2±4.4	70.9±6.7	56.1±16.9	48.5±2.0	**94.9±4.6**
	SHD	7.0±0.8	4.4±1.4	18.3±0.4	12.8±2.9	119.2±8.2	6.2±2.0	9.8±1.7	27.0±8.2	26.3±0.9	**2.0±2.2**
1	SID	41.8±12.2	24.0±2.2	74.6±6.5	60.2±20.0	84.6±35.4	30.6±22.6	34.8±5.8	51.6±25.7	64.6±6.7	**12.4±12.1**
	F1	78.6±4.6	84.1±3.8	40.5±8.2	55.5±12.4	14.1±4.7	80.4±4.4	69.6±4.3	51.9±8.6	46.3±0.5	**93.7±6.0**
Training time		6.7±0.6	36.7±0.7	990.1±93.9	327.7±5.7	343.1±110.6	29.8±1.2	80.9±1.0	33.2±2.6	45.4±2.6	15.8±3.3
SynER1 (d=50)											
	SHD	43.0±4.8	39.8±4.7	26.6±7.7	**5.2±2.8**	795.2±29.5	6.6±3.0	17.0±3.4	96.3±13.9	56.3±6.2	7.2±4.4
0	SID	281.2±114.3	270.4±94.0	173.3±77.6	25.2±7.7	198.6±79.5	**24.8±13.2**	105.8±55.7	262.0±137.3	168.3±50.0	56.6±49.0
	F1	27.5±6.5	34.3±6.1	60.1±12.4	**94.3±2.1**	6.9±1.1	93.2±3.0	77.8±4.6	34.5±3.9	52.4±0.4	91.4±6.2
	SHD	38.0±6.0	32.6±7.3	30.0±4.9	12.2±9.8	806.4±17.8	12.0±3.5	16.8±3.5	98.6±16.7	67.6±9.2	**11.4±1.8**
0.25	SID	248.0±109.4	254.6±94.5	168.6±28.7	**63.2±41.2**	297.8±80.9	79.2±18.6	109.0±51.9	294.0±130.9	195.3±94.4	68.4±18.6
	F1	40.7±10.6	48.5±11.4	58.4±10.2	84.7±10.5	5.7±5.1	84.4±5.4	76.8±4.0	32.5±3.6	47.5±2.4	**85.7±3.3**
	SHD	29.8±4.3	27.4±1.9	25.6±6.9	16.4±7.9	816.6±35.5	12.0±5.5	20.8±4.7	94.0±10.6	80.3±9.5	**11.4±4.4**
0.5	SID	201.4±113.8	186.6±57.8	143.0±8.8	76.0±46.6	231.6±75.7	**65.2±48.2**	134.8±82.8	285.3±175.8	218.6±117.7	72.8±47.5
	F1	58.4±6.7	58.3±6.7	62.1±13.0	79.0±9.6	6.2±0.7	85.6±6.8	72.8±4.3	37.4±2.5	43.1±5.2	**86.5±5.4**
	SHD	26.8±2.6	23.2±3.3	37.6±5.7	26.4±8.8	789.6±26.2	11.8±4.3	19.4±2.3	85.6±7.4	103.0±27.7	**8.2±3.0**
0.75	SID	190.6±86.4	190.8±77.2	267.6±78.8	146.0±58.3	303.6±110.9	53.2±19.5	94.2±40.0	224.0±104.2	165.3±78.2	**35.8±18.2**
	F1	62.7±7.1	64.8±3.7	46.2±10.5	66.0±7.9	5.7±0.7	85.4±5.0	75.9±2.9	41.1±0.8	40.4±4.9	**90.0±3.9**
	SHD	18.2±2.9	18.8±5.5	40.0±1.6	34.0±11.8	782.8±35.2	10.2±4.9	21.2±7.2	94.0±11.4	120.3±6.6	**6.2±2.8**
1	SID	131.2±64.7	139.2±59.1	258.3±98.3	214.4±118.5	344.4±165.1	69.8±42.9	102.8±67.9	152.6±44.7	198.6±99.1	**36.8±21.8**
	F1	77.4±5.6	72.3±7.0	47.2±3.5	58.1±9.9	5.6±0.9	86.6±5.7	74.2±6.4	40.0±1.3	34.6±1.7	**92.3±3.5**
Training time		29.6±6.5	40.1±0.4	3.1k±0.2k	23k±1.8k	322.6±18.7	1.3k±38.4	1.3k±59.9	71.5±2.6	89.1±8.9	319.8±98.8
SynER1 (n=1000)											
	SHD	6.6±1.3	6.2±2.2	3.0±1.4	0.8±1.1	39.0±1.2	0.6±1.2	2.6±1.8	12.6±3.3	11.0±1.6	**0.4±0.8**
0	SID	21.4±12.5	24.4±14.4	7.6±6.2	0.6±1.2	14.6±5.6	0.6±1.2	10.8±11.9	8.6±3.6	9.2±4.2	**0.6±1.2**
	F1	38.4±10.9	36.3±24.2	78.8±9.5	95.9±6.1	22.2±4.2	96.8±6.3	79.6±13.4	48.1±6.3	52.7±7.4	**97.8±4.4**
	SHD	6.6±1.6	6.4±2.1	3.3±1.8	2.6±2.2	39.4±2.5	1.6±1.3	2.6±1.8	9.3±2.6	9.2±2.4	**1.0±1.2**
0.25	SID	20.2±12.4	23.0±15.0	11.3±3.0	7.2±7.7	13.4±6.3	4.2±3.9	12.2±12.1	9.3±4.7	12.0±6.0	**2.2±2.8**
	F1	36.1±22.3	36.0±22.7	68.8±6.3	79.0±15.4	20.6±9.2	86.6±12.4	78.8±14.9	50.7±10.2	52.5±5.1	**93.3±8.2**
	SHD	6.2±1.9	5.0±1.8	4.3±1.2	3.8±2.7	39.4±1.3	2.8±1.6	3.8±0.9	11.3±1.6	12.2±1.6	**2.0±0.8**
0.5	SID	17.0±11.9	15.8±8.8	13.6±2.0	16.6±14.2	14.6±10.6	8.2±8.9	12.2±4.6	11.6±3.0	11.2±8.1	**3.2±1.9**
	F1	44.2±18.8	54.1±17.2	56.5±10.0	67.3±22.4	20.7±15.7	78.6±9.7	64.5±10.8	42.8±3.9	44.8±6.7	**86.8±5.1**
	SHD	4.2±1.7	5.2±2.9	5.3±2.0	1.8±0.4	39.6±2.2	2.2±1.1	3.6±2.1	9.0±2.4	11.0±1.6	**1.4±0.8**
0.75	SID	10.8±8.2	18.0±16.1	16.3±1.6	5.8±4.0	17.6±10.9	7.6±6.6	10.0±6.9	**4.3±4.1**	6.0±2.1	6.2±6.2
	F1	68.9±12.6	56.1±25.6	49.5±12.1	83.5±4.3	19.9±4.9	81.1±11.1	70.3±12.5	60.5±6.9	50.1±7.6	**89.3±6.9**
	SHD	3.0±1.0	4.4±2.6	7.3±4.7	2.2±1.4	40.0±2.1	1.8±0.9	3.2±1.9	10.6±3.6	13.3±2.0	**1.2±0.7**
1	SID	8.4±8.9	17.0±16.5	16.6±6.0	9.6±7.9	18.0±10.4	9.0±6.1	10.6±6.5	**5.0±4.5**	6.0±2.1	5.6±6.6
	F1	79.6±6.7	61.5±23.3	43.5±17.6	77.2±15.6	18.5±7.8	79.9±11.0	71.2±10.5	54.5±11.1	44.7±8.9	**90.2±6.7**
Training time		2.1±0.7	32.2±1.3	471.5±33.7	9.9±0.4	151.1±33.9	4.0±0.9	33.3±2.0	36.5±42.6	34.9±3.7	3.5±0.7
SynER1 (n=5000)											
	SHD	6.4±1.0	5.6±1.2	0.3±0.4	**0.0±0.0**	37.8±1.5	0.6±0.4	2.8±1.8	3.6±1.2	3.0±0.0	0.4±0.8
0	SID	19.4±9.2	18.0±9.8	1.3±1.8	**0.0±0.0**	8.4±7.8	2.4±1.9	10.8±8.7	6.33±3.8	4.3±4.7	0.6±1.2
	F1	41.1±10.7	50.6±5.7	97.7±3.1	**100±0.0**	26.6±5.5	93.2±5.5	74.1±13.5	71.8±13.3	80.6±2.6	96.8±6.3
	SHD	6.2±2.3	5.2±1.6	1.0±0.8	0.6±0.8	39.0±1.0	2.0±0.6	3.0±2.5	3.0±1.6	3.0±2.1	**0.0±0.0**
0.25	SID	17.0±10.7	16.4±11.2	3.6±2.6	2.2±2.7	11.8±7.1	6.0±2.8	11.8±9.7	8.0±4.3	6.0±2.1	**0.0±0.0**
	F1	39.9±28.1	57.3±12.5	91.0±6.3	94.6±6.5	22.2±3.8	81.1±6.7	75.1±19.5	70.2±19.9	81.0±10.7	**100±0.0**
	SHD	5.4±1.8	5.4±2.2	2.0±0.8	1.6±1.0	39.4±0.4	4.6±1.6	4.4±3.6	2.3±1.2	4.3±3.2	**1.4±1.0**
0.5	SID	12.8±7.3	16.4±11.4	6.3±1.2	**2.2±2.7**	16.6±8.0	14.0±6.4	12.4±11.1	5.3±1.6	6.6±1.2	5.4±5.7
	F1	54.0±17.2	54.0±17.6	81.5±7.1	84.1±9.9	20.8±1.7	62.0±14.9	68.1±23.9	78.3±8.2	73.1±10.8	**87.7±7.6**
	SHD	4.2±1.1	4.8±3.0	4.6±1.8	6.2±2.9	39.8±1.7	5.4±2.6	6.0±2.6	5.6±3.3	6.0±2.8	**2.2±2.0**
0.75	SID	11.2±7.6	17.2±16.1	15.0±2.1	20.8±9.5	19.4±11.1	14.2±6.9	17.2±12.1	8.0±4.9	8.0±7.0	**4.4±2.8**
	F1	69.0±6.7	56.8±26.9	54.8±17.9	48.6±25.2	19.2±6.0	62.5±14.5	56.4±22.8	65.4±14.5	64.3±11.6	**84.0±10.8**
	SHD	**3.2±0.9**	4.8±2.4	9.0±3.5	6.2±1.6	40.4±1.3	6.6±3.7	6.2±2.2	5.3±2.8	5.6±3.6	3.4±1.9
1	SID	9.0±8.6	18.4±14.9	20.3±5.1	17.0±8.7	22.0±8.2	17.0±6.9	17.8±9.6	**4.3±1.8**	4.6±0.9	10.2±3.5
	F1	**77.5±6.8**	56.1±20.4	35.3±3.7	48.5±19.3	17.0±4.7	52.3±14.0	48.6±21.5	69.3±8.8	69.2±14.7	74.7±10.6
Training time		3.9±0.6	32.7±0.6	551.1±90.5	55.7±1.1	458.5±194.5	14.3±0.8	42.4±13.5	65.5±33.3	36.6±2.7	15.0±0.4

Experiments – Real-world Datasets

Dataset	Sachs			Syntren		
Metrics	SHD↓	SID↓	F1↑	SHD↓	SID↓	F1↑
NOTEARS	12.0±0.00	46.0±0.00	0.387±0.000	33.9±4.57	192.8±54.73	0.164±0.085
GOLEM	17.0±0.00	44.0±0.00	0.421±0.000	43.7±10.72	177.4±56.55	0.163±0.066
GraNDAG	13.2±0.75	54.0±1.10	0.373±0.064	**26.5±6.45**	155.3±58.11	**0.344±0.104**
CAM	12.0±0.00	55.0±0.00	0.444±0.000	38.0±5.59	178.6±44.56	0.223±0.099
VI-DP-DAG	42.6±1.36	**40.0±5.66**	0.340±0.037	182.6±4.29	**144.3±35.00**	0.069±0.039
SCORE	12.0±0.00	45.0±0.00	0.444±0.000	37.5±4.20	197.1±63.71	0.183±0.091
DiffAN	12.2±0.98	46.2±6.18	0.434±0.078	44.1±8.29	188.7±55.16	0.191±0.095
DAGuerreotype	17.9±0.54	51.4±0.49	0.118±0.034	87.9±9.60	157.7±48.90	0.125±0.047
CaPS	**11.0±0.00**	42.0±0.00	**0.500±0.000**	37.2±5.04	178.9±55.58	0.230±0.072
w/o Theorem 1	17.0±3.50	54.0±3.40	0.257±0.061	51.6±8.82	180.0±66.80	0.218±0.090
w/o Parent Score	12.0±0.00	45.0±0.00	0.444±0.000	34.8±3.37	188.0±57.58	0.222±0.083

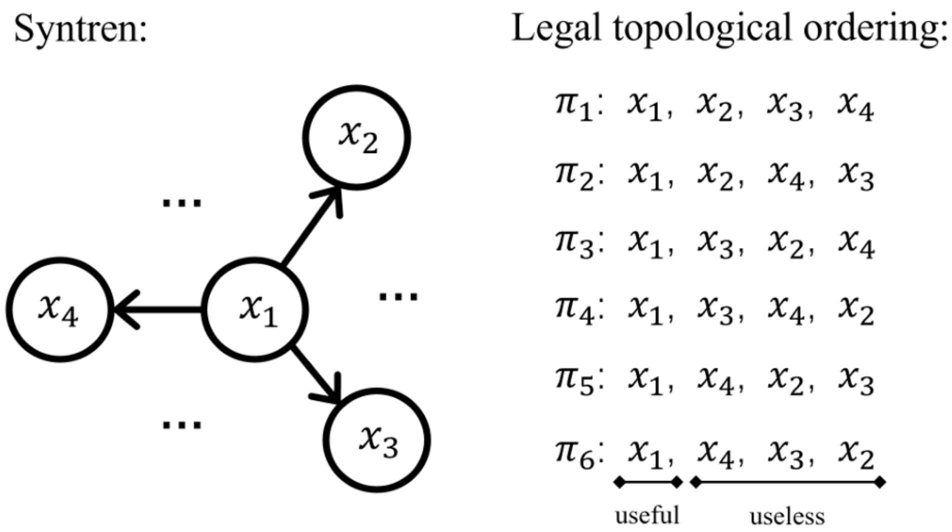


Figure 4: Example of the Syntren dataset.

- In real-world datasets, CaPS achieves the **highest SHD and F1 scores on Sachs** and the **second best F1 on Syntren.**
- The pattern of Syntren is not friendly to ordering-based methods, since it is **a special dataset containing many star networks**. However, CaPS achieves the best performance compared to other ordering-based methods.

Experiments – Acceleration & Visualization

- The acceleration percentage of pre-pruning becomes **more significant** when the number of nodes d grows.

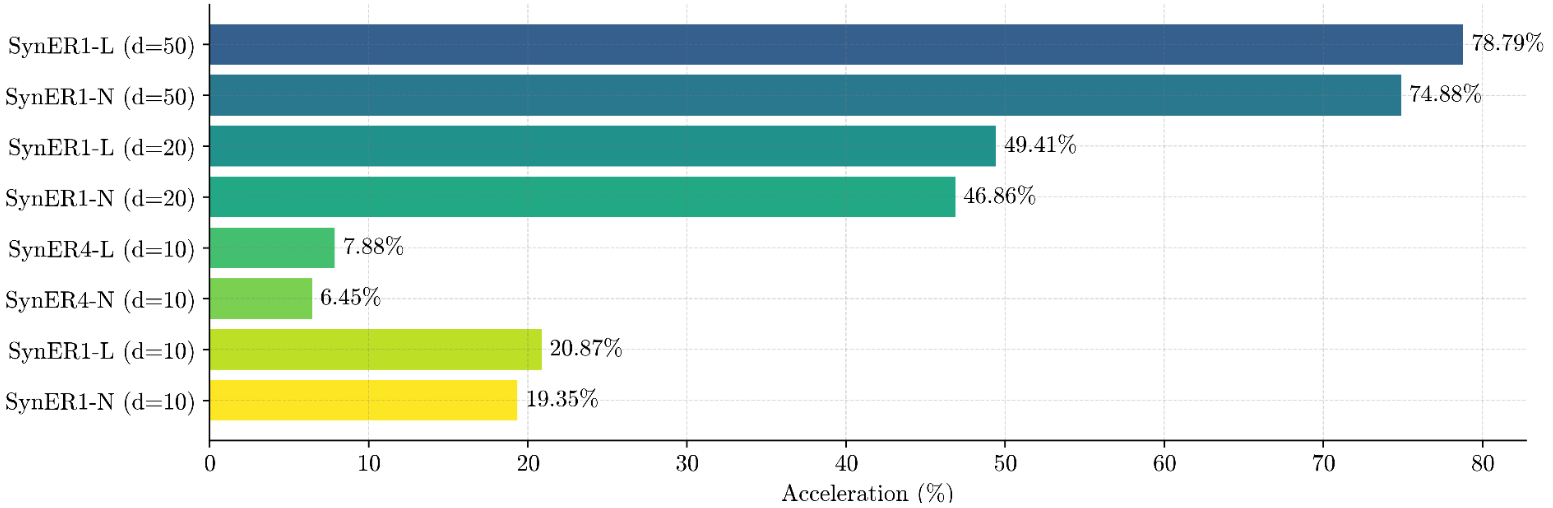


Figure 5: Percentage of acceleration using pre-pruning.

- The parent score captures most of the ground-truth edges and the estimated weights are **similar to the actual values**.

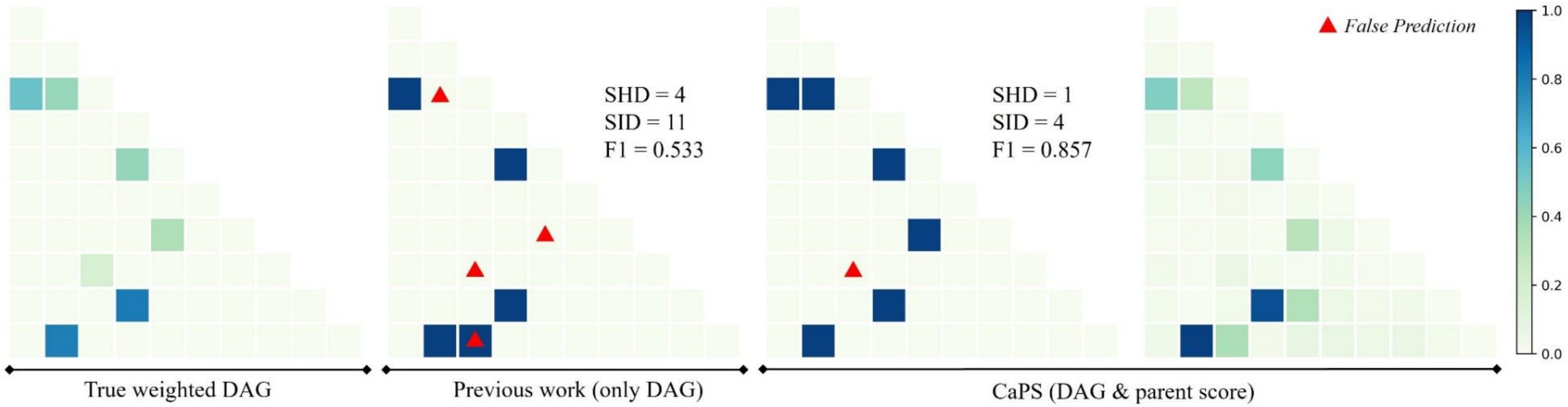


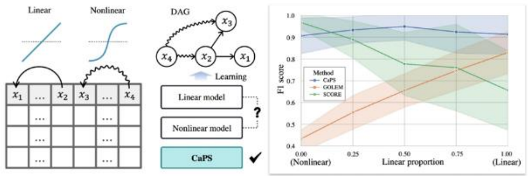
Figure 6: Visualization on SynER1 dataset. Darker colors indicate stronger causal effects.

Ordering-Based Causal Discovery for Linear and Nonlinear Relations

Zhuopeng Xu, Yujie Li, Cheng Liu, Ning Gui*
Central South University

Linear and Nonlinear Challenge in Causal Discovery

Causal discovery uncovers causal relationships within data by modeling a Directed Acyclic Graph (DAG) connecting various variables. Existing approaches normally limit their discussions to distributions with either pure linear or pure nonlinear relations. However, **real-world data often contain both types of causal relations** and run against their basic assumptions.



Real-world data (linear / nonlinear / mixed)

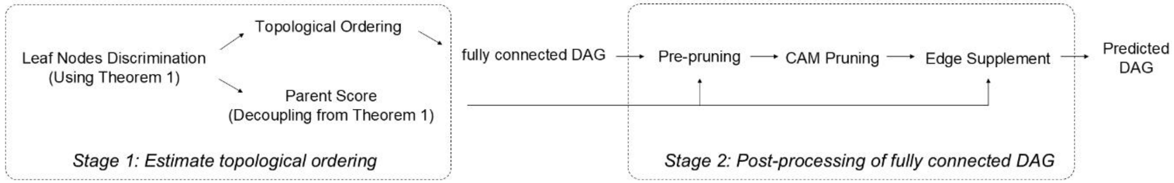
- *Linear model, e.g., GOLEM* → *Nonlinear* ↑ → *Performance* ↓
- *Nonlinear model, e.g., SCORE* → *Linear* ↑ → *Performance* ↓
- *Our model, CaPS* → *Consistently work well*

We don't know whether the real-world data is linear or nonlinear.

We need a method that **works well in both linear and nonlinear and most possibly mixed** cases.

CaPS: Causal Discovery with Parent Score

Overview of CaPS:



A novel ordering criterion for leaf nodes discrimination.

Sufficient conditions for identifiability without any assumption of causal relations.

(i) Non-decreasing variance of noises. (ii) Non-weak causal effect.

Conditions *(i)* and *(ii)* are two different identifiable scenarios, and CaPS only needs **one of them** to be satisfied. (see Assumption 1 for more details)

Theorem 1. *Let* $s(x) = \nabla \log p(x)$ *be the score and let* $diag(\cdot)$ *be the diagonal elements of the matrix. For any* x_j *in the causal graph* $\mathcal{G}$*:*

$$j = \text{argmax}(diag(\mathbb{E}\left[\frac{\partial s(x)}{\partial x}\right])) \Rightarrow x_j \text{ is a leaf node}$$

Under the sufficient conditions *(i)* **or** *(ii)*, the topological ordering is identifiable by iteratively eliminating the current leaf node using Theorem 1.

A new metric to approximate the average causal effect.

A new metric, parent score, is introduced to **reflects the strength of the average causal effect** of a given parent. This metric can be obtained directly by decoupling from Theorem 1 **without any additional computational complexity**.

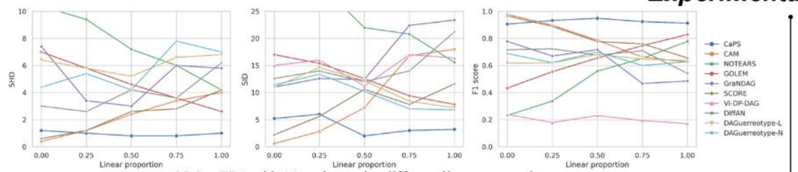
Parent Score.

$$\mathcal{P}_{i,j} = \begin{cases} \frac{1}{\sigma_i^2}\mathbb{E}\left[\left(\frac{\partial f_i}{\partial x_j}(pa_i(x))\right)^2\right], & x_j \in pa_i(x) \\ 0, & x_j \notin pa\_i\ (x) \end{cases}$$

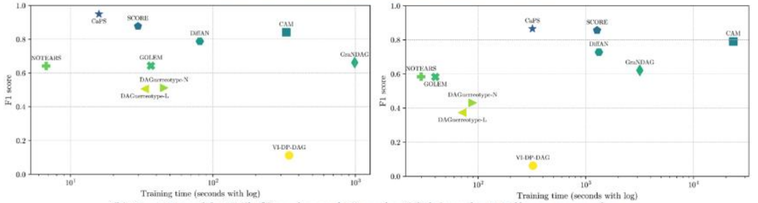
speed up → **Pre-pruning.** Remove the low-confidence edges and reduce the searching space.

rectify → **Edge Supplement.** Use high-confidence parents to supplement the edge.

Experimental results



(a) SynER1 with 10 nodes under different linear proportion



(b) SynER1 with 20 (left) nodes and 50 nodes (right) under 0.5 linear proportion

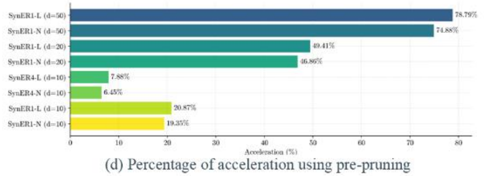
In our manuscript, we discussed the performance of CaPS under different scenarios, e.g.

1. *Different linear proportion*
2. *Different sparsity*
3. *Different DAG type*
4. *Larger-scale sample size*
5. *Larger-scale node size*
6. *Beyond our assumptions*
7. *Order divergence*
8. *Real-world datasets*

CaPS can achieve consistently good performance under each setting while its time cost is competitive.

Dataset	Sachs			Syntren		
Metrics	SHD↓	SID↓	F1↑	SHD↓	SID↓	F1↑
NOTEARS	12.0±0.00	46.0±0.00	0.387±0.000	33.9±4.57	192.8±54.73	0.164±0.085
GOLEM	17.0±0.00	44.0±0.00	0.421±0.000	43.7±10.72	177.4±56.55	0.163±0.066
GraNDAG	13.2±0.75	54.0±1.10	0.373±0.064	**26.5±6.45**	155.3±58.11	**0.344±0.104**
CAM	12.0±0.00	55.0±0.00	0.444±0.000	38.0±5.59	178.6±44.56	0.223±0.099
VI-DP-DAG	42.6±1.36	**40.0±5.66**	0.340±0.037	182.6±4.29	**144.3±35.00**	0.069±0.039
SCORE	12.0±0.00	45.0±0.00	0.444±0.000	37.5±4.20	197.1±63.71	0.183±0.091
DiffAN	12.2±0.98	46.2±6.18	0.434±0.078	44.1±8.29	188.7±55.16	0.191±0.095
DAGuerreotype	17.9±0.54	51.4±0.49	0.118±0.034	87.9±9.60	157.7±48.90	0.125±0.047
CaPS	**11.0±0.00**	42.0±0.00	**0.500±0.000**	37.2±5.04	178.9±55.58	0.230±0.072
w/o Theorem 1	17.0±3.50	54.0±3.40	0.257±0.061	51.6±8.82	180.0±66.80	0.218±0.090
w/o Parent Score	12.0±0.00	45.0±0.00	0.444±0.000	34.8±3.37	188.0±57.58	0.222±0.083

(c) Results of real-world datasets



(d) Percentage of acceleration using pre-pruning

QR code

Code of CaPS

Paper of CaPS

Poster Session: Wed 11 Dec 4:30 p.m. PST — 7:30 p.m. PST

Email: *xuzhuopeng@csu.edu.cn, ninggui@gmail.com*

CENTRAL SOUTH UNIVERSITY
NEURAL INFORMATION PROCESSING SYSTEMS

Thank you!