Learning DAGs from Data with Few Root Causes

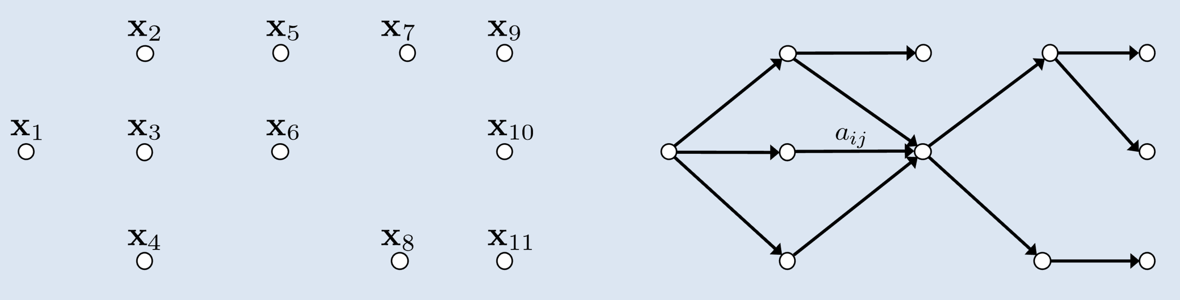
Panagiotis Misiakos, Chris Wendler and Markus Püschel

Computer Science
ETH zürich

Goal: DAG Learning

Given: Data $\mathbf{X} = \{\mathbf{x}_i\}_{1 \le i \le d}$ associated with the nodes of an unknown weighted DAG

Goal: Learn the weighted DAG $\mathbf{A} = \{a_{ij}\}_{1 \le i < j \le d}$



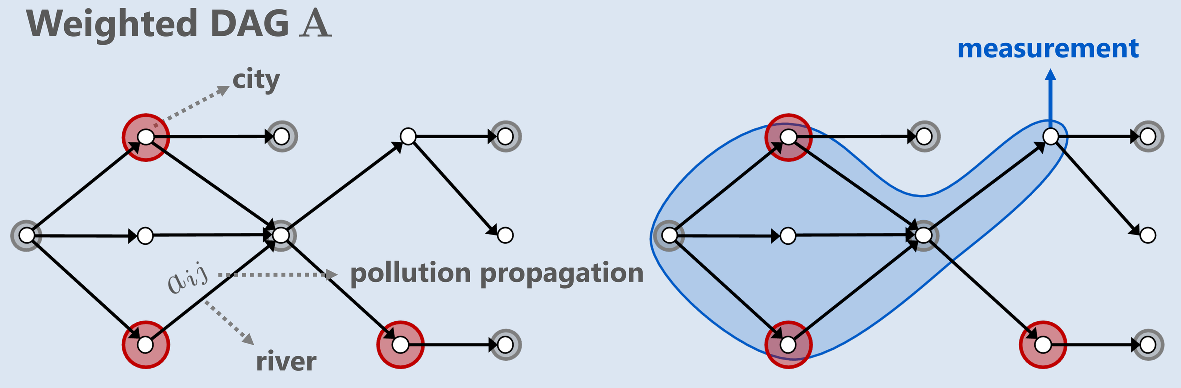
Active research area! [Vowels et. al., 2021], NOTEARS [Zheng et. al., 2018], GOLEM [Ng. et. al., 2020], GraN-DAG [Lachapelle et. al., 2019], [Chevalley et. al., 2023]

Novel assumptions: Few root causes

1. Data $\mathbf{X}$ generated from **few events** upstream
2. $\mathbf{X}$ subject to **measurement noise**

River network example

Weighted DAG $\mathbf{A}$



Few cities pollute $\mathbf{C}$
Negligible pollution by others $\mathbf{N}_c$

Accumulated pollution $\mathbf{X}$
Measurement noise $\mathbf{N}_x$

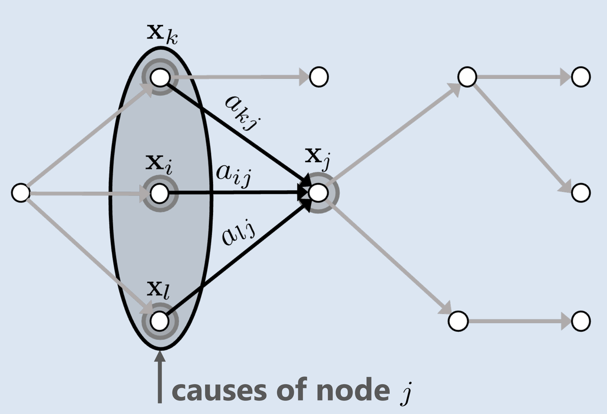
Metaphor for data generated by few events

Data generation model

Linear SEM (structural equation model)

[Shimizu et. al., 2006]

Data are linear combination of the parent's values (causes)



$$\mathbf{x}_j = \sum_{i \in \text{causes}(j)} a_{ij}\mathbf{x}_i + \mathbf{n}_j$$

noise

Solving the recurrence

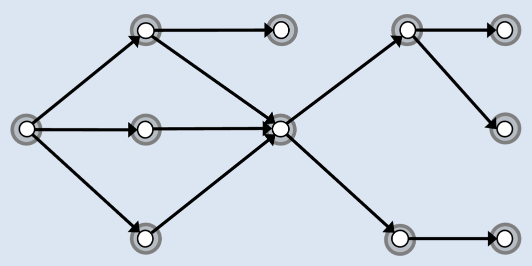
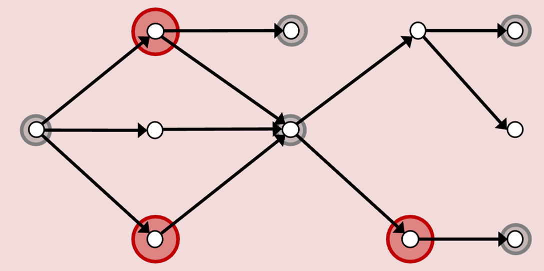
$$\mathbf{X} = \mathbf{X}\mathbf{A} + \mathbf{N} \Leftrightarrow \mathbf{X} = \mathbf{N}\left(\mathbf{I} - \mathbf{A}\right)^{-1}$$
$$= \mathbf{N}\left(\mathbf{I} + \mathbf{A} + \ldots + \mathbf{A}^{(d-1)}\right)$$
$$= \mathbf{N}\left(\mathbf{I} + \overline{\mathbf{A}}\right) \quad \text{Transitive closure}$$

Output — **Input = root causes**

Linear SEM = ***Linear transformation with input $\mathbf{N}$***

Our contribution

New data generation assumptions

Prior work	Ours: Few root causes
Data generation model	
$\mathbf{X} = \mathbf{N}\left(\mathbf{I} + \overline{\mathbf{A}}\right)$	$\mathbf{X} = \left(\mathbf{C} + \mathbf{N}_c\right)\left(\mathbf{I} + \overline{\mathbf{A}}\right) + \mathbf{N}_x$
Linear SEM	Linear SEM with **few root causes** and **measurement noise**
Input (root causes)	
	
Input: $\mathbf{N}$	Input: $\mathbf{C} + \mathbf{N}_c$
Random and i.i.d.	Approximately sparse
	$\mathbf{C}$ is sparse with varying support
$\mathbf{N}$: low magnitude noise	**$\mathbf{N}_c$: low magnitude noise**
Weighted DAG	
No weight constraint in $\mathbf{A}$	$\mathbf{A}$ has weights in $[0, 1]$
$\mathbf{A}$ is sparse	$\mathbf{A}$ is sparse
Data/Measurements $\mathbf{X}$	
Assumed to be exact	**Subject to measurement noise $\mathbf{N}_x$**

Learning the DAG

Theoretical Guarantees

Lemma: Given $\mathbf{N}_x = 0$ the DAG is **identifiable** from data with **few root causes**.

Proof is based on identifiability from Linear non-Gaussian SEM [Shimizu et. al., 2006].

Theorem: Given $\mathbf{N}_c = \mathbf{N}_x = 0$ and enough data the true DAG $\mathbf{A}$ is (with high probability) the **minimizer** of:

$$\min_{\mathbf{A} \in \mathbb{R}^{d \times d}} \left\| \mathbf{X}\left(\mathbf{I} + \overline{\mathbf{A}}\right)^{-1} \right\|_0 \quad \text{s.t. } \mathbf{A} \text{ is acyclic}$$

Proof in our paper.

Learning the DAG by continuous relaxation

In practice $\mathbf{N}_c \neq 0,\ \mathbf{N}_x \neq 0$ and we apply the L^1 norm.

$$\min_{\mathbf{A} \in \mathbb{R}^{d \times d}} \left\| \mathbf{X}\left(\mathbf{I} + \overline{\mathbf{A}}\right)^{-1} \right\|_1 + \lambda \|\mathbf{A}\|_1 \quad \text{s.t. } \operatorname{tr}\left(e^{\mathbf{A} \odot \mathbf{A}}\right) = d$$

Few root causes — Sparse DAG — Acyclicity constraint NOTEARS [Zheng et. al., 2018]

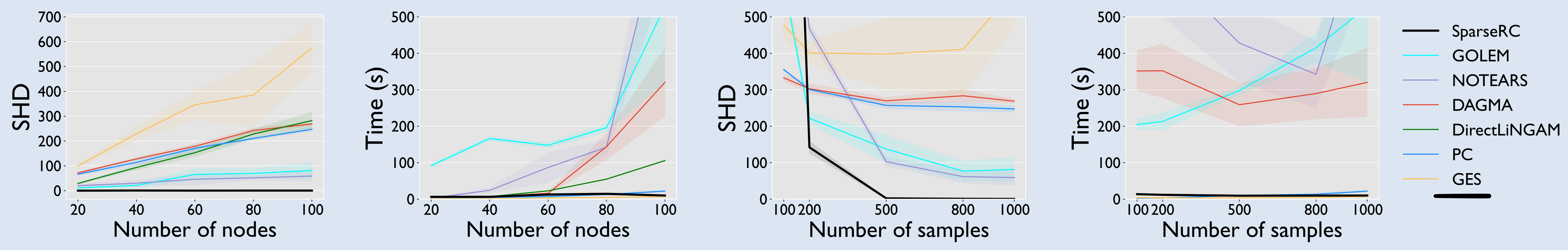
Experiments

Default experiment: few root causes assumption fullfiled by construction

- $\mathbf{C}$ is multivariate Bernoulli with $p = 0.1$ and weights from $\mathcal{U}(0, 1)$
- Both root causes $\mathbf{N}_c$ and measurement noise $\mathbf{N}_x$ have low std. $\sigma = 0.01$
- Random Erdös-Renyi graph (or other) transfomed into DAG
- Average degree $= 4$ and weights from $[-0.9, -0.1] \cup [0.1, 0.9]$

Evaluation metrics: **SHD** (structural Hamming distance), **runtime**

SparseRC is best for all nodes for more than 500 samples



Also benchmarked but not competitive
DAGMA
DirectLiNGAM
PC
GES
LiNGAM
CAM
DAG-NoCurl
fGES
sortnregress
MMHC

Different variation of the default experiment (100 nodes, 400 edges, 1000 samples)

Excellent reconstruction when assumptions are fullfilled — **10-50x faster**

	Hyperparameter	Default	Change	Reconstruction quality (SHD): SparseRC (ours)	GOLEM	NOTEARS	Runtime [seconds]: SparseRC (ours)	GOLEM	NOTEARS
1	Default settings			$\mathbf{0.6 \pm 0.8}$	82 ± 34	59 ± 22	$\mathbf{10 \pm 1.8}$	529 ± 210	796 ± 185
2	Graph type	Erdös-Renyi	Scale-free	$\mathbf{2.2 \pm 1.5}$	34 ± 9.0	28 ± 9.5	$\mathbf{11 \pm 1.1}$	460 ± 184	180 ± 7.2
3	$\mathbf{N}_c, \mathbf{N}_x$ distribution	Gaussian	Gumbel	$\mathbf{1.4 \pm 1.0}$	87 ± 44	59 ± 17	$\mathbf{8.2 \pm 0.7}$	349 ± 125	251 ± 48
4	Edges / Vertices	4	10	$\mathbf{46 \pm 7.5}$	212 ± 70	243 ± 26	$\mathbf{14 \pm 1.0}$	347 ± 121	471 ± 82
5	Standardization	No	Yes	624 ± 48	failure	failure	$\mathbf{13 \pm 0.7}$	194 ± 9.6	679 ± 72
6	Larger weights in $\mathbf{A}$	$(0.1, 0.9)$	$(0.5, 2)$	failure	96 ± 25	$\mathbf{92 \pm 14}$	$\mathbf{11 \pm 1.9}$	326 ± 145	781 ± 76
7	$\mathbf{N}_c, \mathbf{N}_x$ deviation	$\sigma = 0.01$	$\sigma = 0.1$	504 ± 19	$\mathbf{98 \pm 14}$	199 ± 12	$\mathbf{8.4 \pm 0.6}$	431 ± 177	2834 ± 228
8	Dense root causes $\mathbf{C}$	$p = 0.1$	$p = 0.5$	1221 ± 33	$\mathbf{29 \pm 2.5}$	126 ± 32	$\mathbf{8.7 \pm 0.7}$	309 ± 63	433 ± 53
9	Samples	$n = 1000$	$n = 100$	2063 ± 92	failure	failure	$\mathbf{9.1 \pm 0.7}$	334 ± 121	427 ± 35
10	Fixed support	No	Yes	failure	failure	failure	$\mathbf{15 \pm 2.0}$	360 ± 142	669 ± 386

Assumptions deteriorate

5-8: deteriorate the sparsity in root causes
9: Low number of samples
10: violates varying support

Scaling to larger DAGs

Reconstruction quality (SHD)

Nodes d, samples n	SparseRC	NOTEARS	GOLEM
$d = 200,\ n = 500$	22	155	281
$d = 500,\ n = 1000$	27	245	574
$d = 1000,\ n = 5000$	26	282	699
$d = 2000,\ n = 10000$	50	489	time-out
$d = 3000,\ n = 10000$	134	time-out	time-out

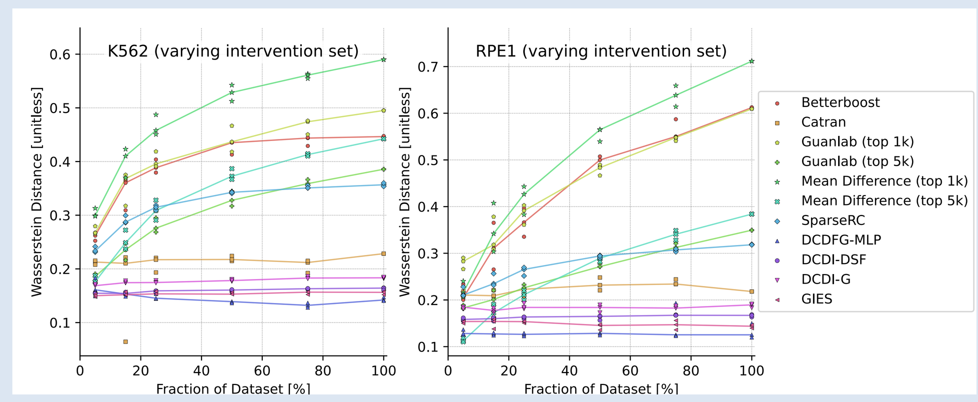
Excellent reconstruction

Real Data [Sachs et. al., 2005]

	SHD ↓	SID ↓	Total edges
SparseRC	15	45	16
NOTEARS	**11**	44	15
GOLEM	21	**43**	19

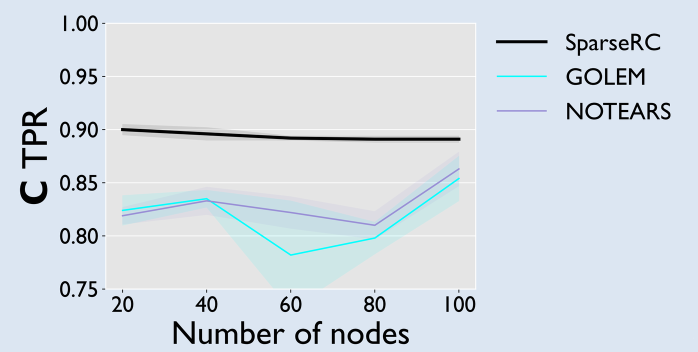
Competitive performance on real data

CausalBench Challenge [Chevalley et. al., 2023]

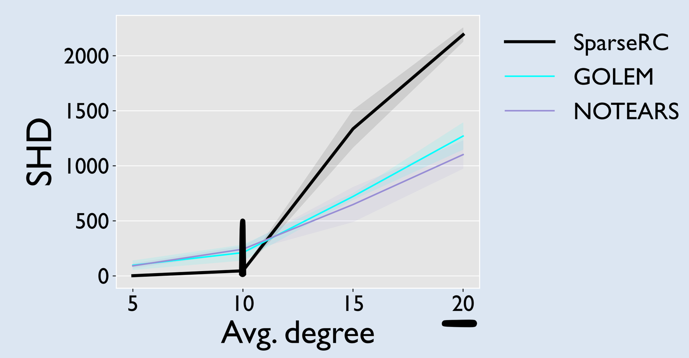


SparseRC ranked 3rd in the CausalBench challenge at ICLR 2023

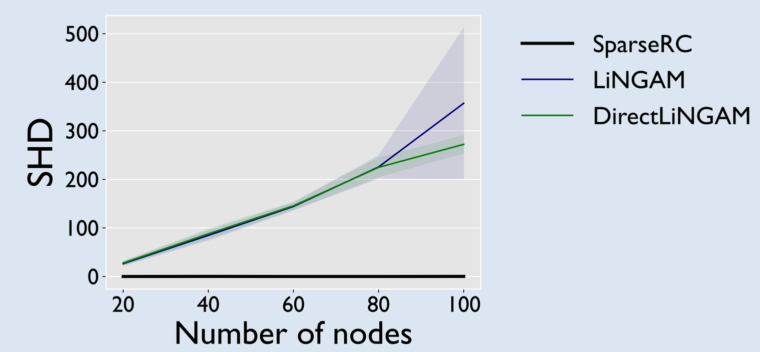
Root causes recovery



Denser DAGs



Zero measurement noise



*Checkout our **github repo***