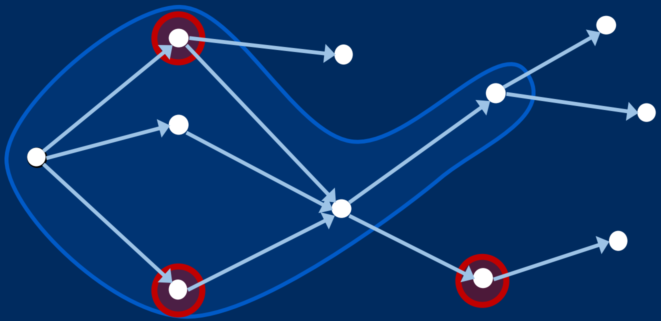
Learning DAGs from Data with Few Root Causes

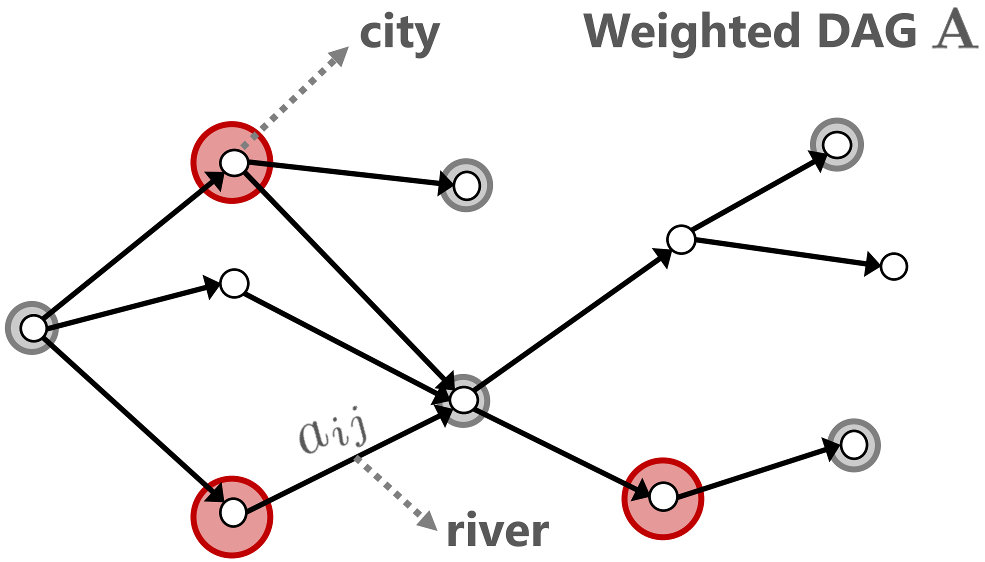


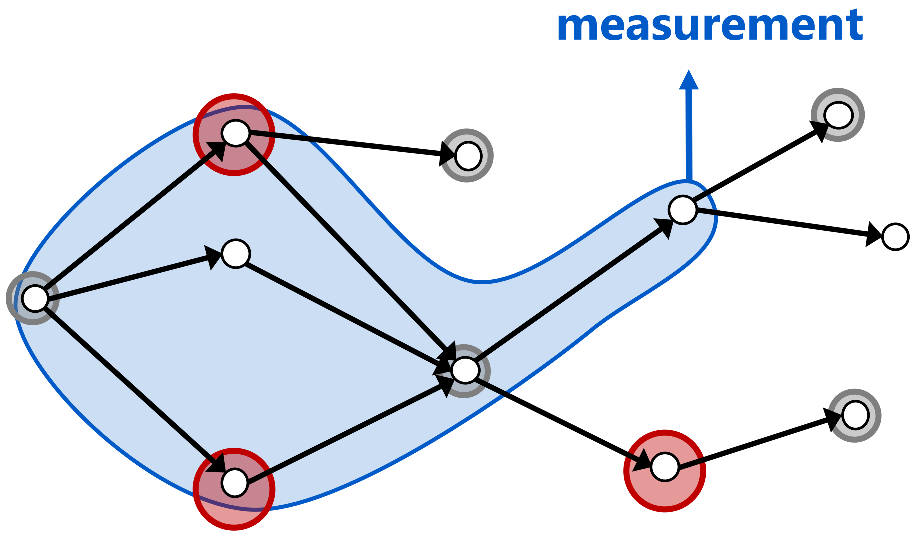
Panagiotis Misiakos, Chris Wendler and Markus Püschel

Department of Computer Science
ETH zürich



Motivation: Pollution on a river network





Few cities pollute $\mathbf{C}$

Negligible pollution by others N_c

Measurement of accumulated pollution $\mathbf{X}$

Measurement noise N_x

Follows a linear structural equation model (SEM)

accumulation

$$\mathbf{X} = (\mathbf{C} + \mathrm{N}_c)\left(\mathbf{I} + \mathbf{A} + \ldots + \mathbf{A}^{d-1}\right) + \mathrm{N}_x \Leftrightarrow$$

$$\mathbf{X} = \mathbf{X}\mathbf{A} + (\mathbf{C} + \mathbf{N}_c) + \mathbf{N}_x\,(\mathbf{I} - \mathbf{A})$$

Goal: Given X, learn the DAG $\mathbf{A}$

Goal: Learn DAG from Data with Few Root Causes

i.i.d. noise

Standard Linear SEM: $\mathbf{X} = \mathbf{X}\mathbf{A} + \mathbf{N}$ no measurement noise

Our model: $\mathbf{X} = \mathbf{X}\mathbf{A} + \underbrace{(\mathbf{C} + \mathbf{N}_c)} + \mathbf{N}_x(\mathbf{I} - \mathbf{A})$

approximately sparse (root causes)

measurement noise

Our work: SparseRC = Finding the DAG $\mathbf{A}$ by solving

$$\min_{\mathbf{A}\in\mathbb{R}^{d\times d}} \left\| \mathbf{X}\left(\mathbf{I} + \overline{\mathbf{A}}\right)^{-1} \right\|_1 + \lambda \|\mathbf{A}\|_1 \quad \text{s.t. } \operatorname{tr}\left(e^{\mathbf{A}\odot\mathbf{A}}\right) = 0$$

Few root causes

Sparse DAG

Acyclicity constraint NOTEARS [Zheng et. al., 2018]

Excellent reconstruction if assumptions are fulfilled **+ 10-50x faster**

Reconstruction (SHD)

	Hyperparameter	Default	Change	SparseRC (ours)	GOLEM	NOTEARS
1	Default settings			**0.6 ± 0.8**	82 ± 34	59 ± 22
2	Graph type	Erdös-Renyi	Scale-free	**2.2 ± 1.5**	34 ± 9.0	28 ± 9.5
3	$\mathbf{N}_c, \mathbf{N}_x$ distribution	Gaussian	Gumbel	**1.4 ± 1.0**	87 ± 44	59 ± 17
4	Edges / Vertices	4	10	**46 ± 7.5**	212 ± 70	243 ± 26
5	Standardization	No	Yes	624 ± 48	failure	failure
6	Larger weights in $\mathbf{A}$	$(0.1, 0.9)$	$(0.5, 2)$	failure	96 ± 25	**92 ± 14**
7	$\mathbf{N}_c, \mathbf{N}_x$ deviation	$\sigma = 0.01$	$\sigma = 0.1$	504 ± 19	**98 ± 14**	199 ± 12
8	Dense root causes $\mathbf{C}$	$p = 0.1$	$p = 0.5$	1221 ± 33	**29 ± 2.5**	126 ± 32
9	Samples	$n = 1000$	$n = 100$	2063 ± 92	failure	failure
10	Fixed support	No	Yes	failure	failure	failure

Also benchmarked but not competitive

DAGMA
DirectLiNGAM
PC
GES
LiNGAM
CAM
DAG-NoCurl
fGES
sortnregress
MMHC

Scalability to larger DAGs (~1000s of nodes)

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

SparseRC effectively reconstructs the weights

SparseRC ranked 3rd in the CausalBench challenge at ICLR 2023 [Chevalley et. al., 2023]