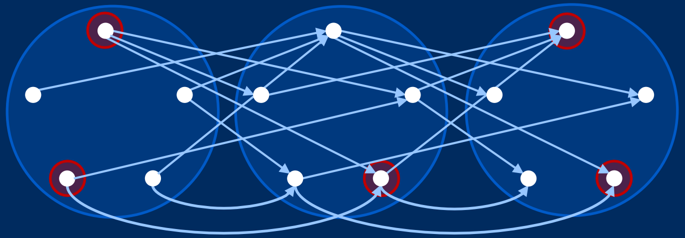
Learning Signals and Graphs from Time-series Graph Data with Few Causes

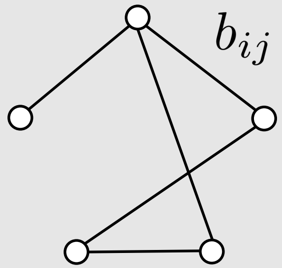
Panagiotis Misiakos, Vedran Mihal and Markus Püschel

Department of Computer Science
ETH zürich



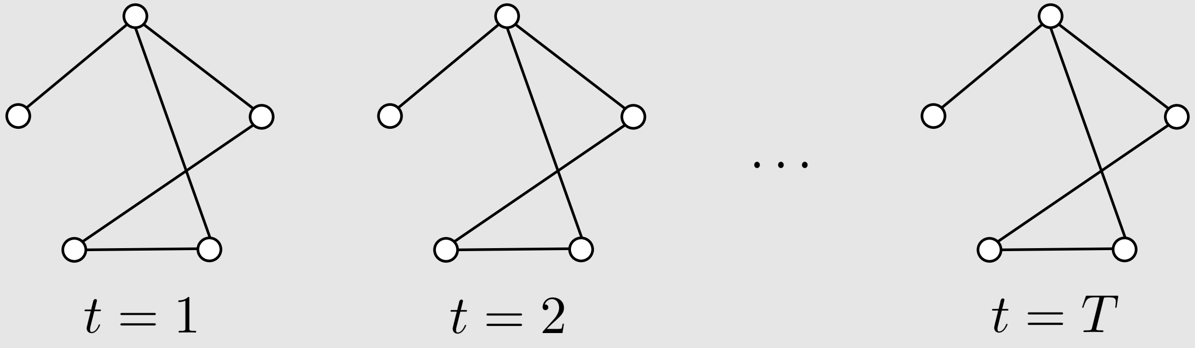
Time-series data on graphs

Time-constant (stationary) graph



Graph $\mathcal{G} = (V, \mathbf{B})$
d nodes

Graph signal: $\mathbf{x} = (x_1, ..., x_d)^T$



Time-series graph data: $\mathbf{x}(t), \ \ t = 1, 2, ..., T$

Time series: $\overline{\mathbf{x}} = \big(\mathbf{x}(1), \mathbf{x}(2), \ ..., \ \mathbf{x}(T)\big)^T$

Collected in matrix: $\mathbf{X} = \big(\overline{\mathbf{x}}_1, \ \overline{\mathbf{x}}_2, \ ..., \ \overline{\mathbf{x}}_n\big)$

Example: S&P stock time series

Nodes

Date	INTC	MSFT	NVDA	GOOG	AAPL	**Graph signal**
2023-01-03	[redacted]	239.6	143.1	89.7	[redacted]	$\mathbf{x}(1)$
2023-01-04	27.7	[redacted]	[redacted]	[redacted]	[redacted]	$\mathbf{x}(2)$
2023-01-05	27.6	[redacted]	142.6	86.8	125.0	$\mathbf{x}(3)$
2023-01-06	[redacted]	224.9	[redacted]	[redacted]	129.6	$\mathbf{x}(4)$
2023-01-09	29.3	[redacted]	156.3	[redacted]	[redacted]	$\mathbf{x}(5)$
$\vdots$	$\vdots$					$\vdots$

Close values

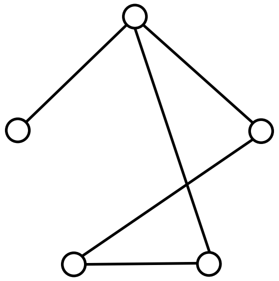
Daily time steps

?

1. Signal learning

?

2. Graph learning



relations between stocks

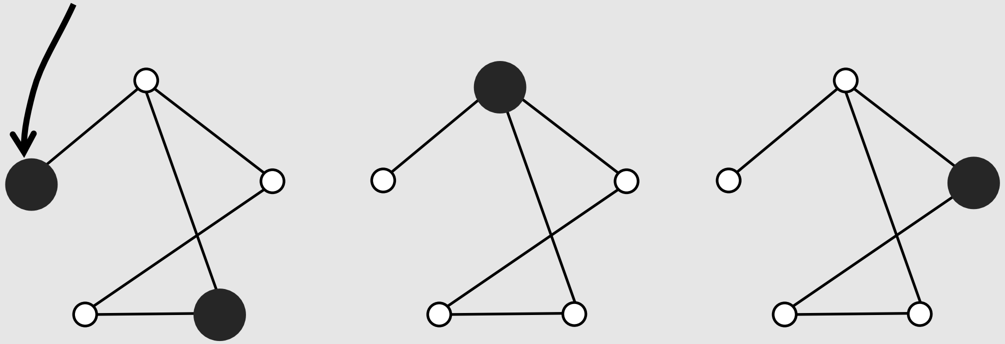
Time series: $\bar{\mathbf{x}} = (\mathbf{x}(1), \mathbf{x}(2), ...) = (26.7, 239.6, 143.1, 89.7, 125.1,\ 27.7, 229.1, 147.5, 88.7, 126.4,\ ...)^T$

Collected time series: e.g. for 5 years $\mathbf{X} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3, \bar{\mathbf{x}}_4, \bar{\mathbf{x}}_5)$

Contributions

1. Signal learning

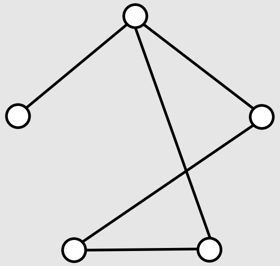
Reconstruct entire time series $\overline{\mathbf{x}}$ from k samples



Sampling and reconstruction: smoothness or Fourier sparsity [Xiao et. al., 2023]

2. Graph learning

Learn the graph $\mathbf{B}$ from n time series $\mathbf{X} = (\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, ..., \overline{\mathbf{x}}_n)$



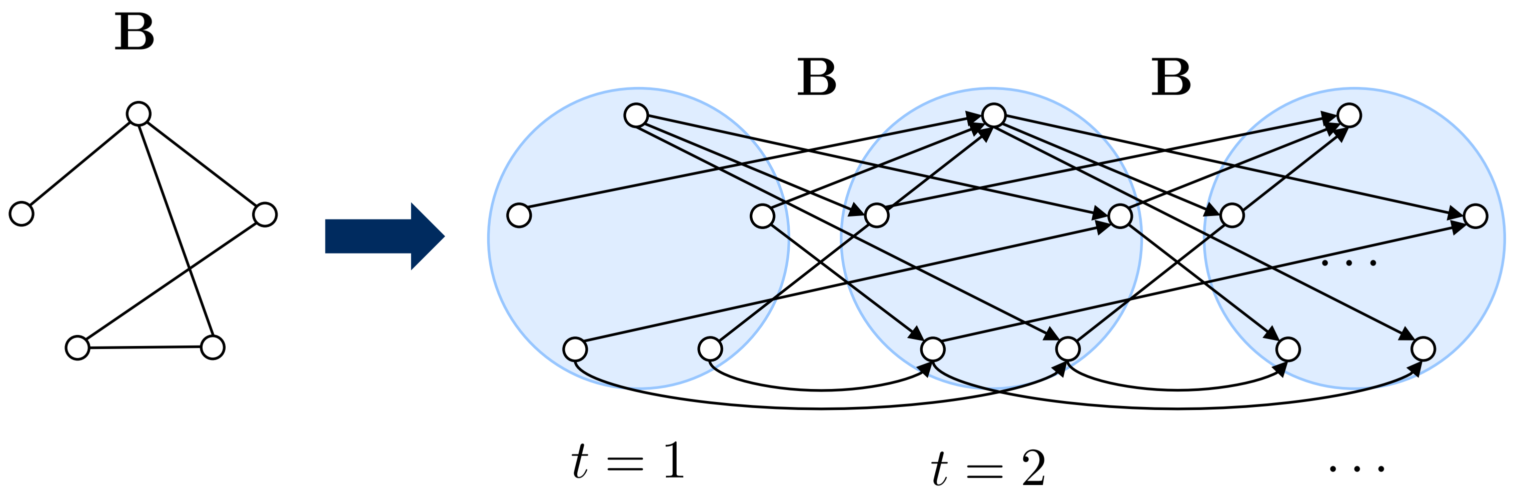
$\mathbf{B} = ?$

Graph learning: smoothness or Fourier sparsity [Dong et. al., 2016, Humbert et. al., 2021]
Learning DAGs (directed acyclic graphs) [Zheng et. al., 2018]

Key idea: Import methods from causal inference on DAGs

Time series of Graphs = Directed Acyclic Graph (DAG)

[Kim and Anderson, 2012]



Adjacency matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \boxed{\mathbf{B}} & 0 & \dots & 0 & 0 \\ \vdots & \mathbf{B} & \ddots & \vdots & \\ & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \dots & \mathbf{B} & 0 \end{pmatrix}$$

Is a DAG!

structural shocks or “causes”

Time-series model:
$$\begin{cases} \mathbf{x}(1) = \mathbf{c}(1) \\ \mathbf{x}(t) = \mathbf{B}\mathbf{x}(t-1) + \mathbf{c}(t), \; t = 2, \ldots, T \end{cases} \quad \Leftrightarrow \quad \overline{\mathbf{x}} = \mathbf{A}\overline{\mathbf{x}} + \overline{\mathbf{c}}$$

Vector autoregression
(signal processing – econometrics)

Linear structural equation model
(machine learning with DAGs)

Idea: We learn B (graph) by learning the entire A (DAG) and then extracting B

Few Causes

Time-series model

$$\overline{\mathbf{x}} = \mathbf{A}\overline{\mathbf{x}} + \overline{\mathbf{c}} \Leftrightarrow \overline{\mathbf{x}} = \left(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \ldots + \mathbf{A}^{T-1}\right)\overline{\mathbf{c}} = \mathbf{W}\overline{\mathbf{c}}$$

time series

causes

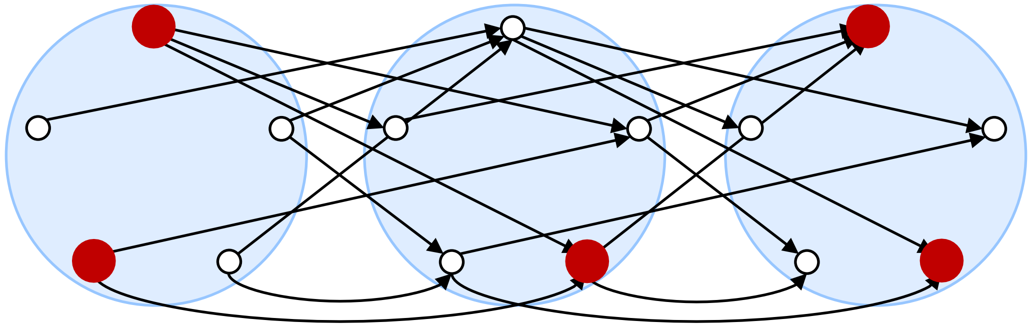
Causal basis matrix

$$\mathbf{W} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{B}^2 & \mathbf{B} & \mathbf{I} & \ddots & \vdots & \\ \mathbf{B}^3 & \mathbf{B}^2 & \mathbf{B} & \ddots & \mathbf{0} & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{I} & \mathbf{0} \\ \mathbf{B}^{(T-1)} & \mathbf{B}^{(T-2)} & \ldots & \mathbf{B}^2 & \mathbf{B} & \mathbf{I} \end{pmatrix}$$

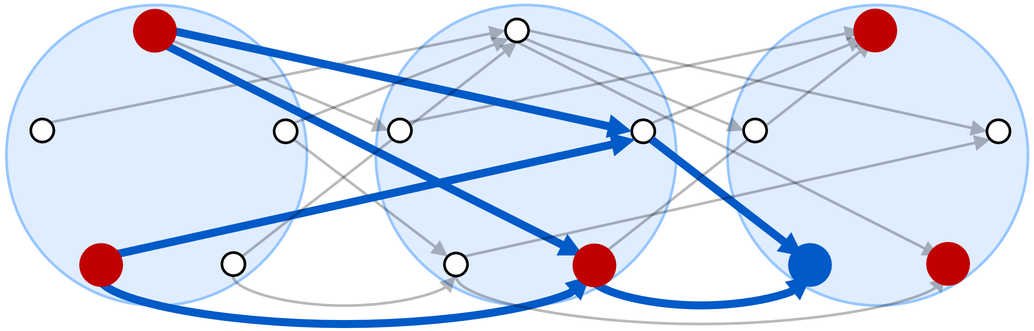
Assumption: $\overline{\mathbf{x}}$ is determined by (approximately) few causes = $\overline{\mathbf{c}}$ is sparse

... *is a form of Fourier sparsity* [Seifert et. al., 2021]

S&P stock example



Causes $\overline{\mathbf{c}}$: financial triggers (new product introduction, change in management)



Time series $\overline{\mathbf{x}} = \mathbf{W}\overline{\mathbf{c}}$: stock value

Proposed methodologies

1. Signal learning

Reconstruct entire time series $\overline{\mathbf{x}}$ from K samples $\mathbf{\Phi}_K\overline{\mathbf{x}}$

Solution: Lasso regression

$$\hat{\overline{\mathbf{x}}} = \mathbf{W} \cdot \arg\min_{\overline{\mathbf{c}}} \frac{1}{2K} \|\mathbf{\Phi}_K \overline{\mathbf{x}} - \mathbf{\Phi}_K \mathbf{W}\overline{\mathbf{c}}\|_2^2 + \lambda \|\overline{\mathbf{c}}\|_1$$

Few causes

$\mathbf{W}$ is our causal (Fourier) basis

Can be replaced by other notions of Fourier bases (adjacency, Laplacian)

2. Graph learning

Learn the weighted graph $\mathbf{B}$ from n time series $\mathbf{X} = \left(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \ldots, \overline{\mathbf{x}}_n\right)$

Solution: DAG Learning

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}\in\mathbb{R}^{dT\times dT}} \|\mathbf{X} - \mathbf{X}\mathbf{A}\|_1 + \lambda\|\mathbf{A} - \mathbf{P}\mathbf{A}\mathbf{Q}\|_2$$

Few causes **Block diagonal**

$$\text{s.t.} \quad \operatorname{tr}\left(e^{\mathbf{A}\odot\mathbf{A}}\right) = 0$$

Acyclicity constraint
[Zheng et. al., 2018]

From $\hat{\mathbf{A}}$ we extract $\hat{\mathbf{B}}$ as the first block

Problem 1: Learning signal from samples

Settings

20 nodes, 50 time steps = Unrolled DAG with 1000 nodes

10 time series, 5% sparsity in the causes

Comparison

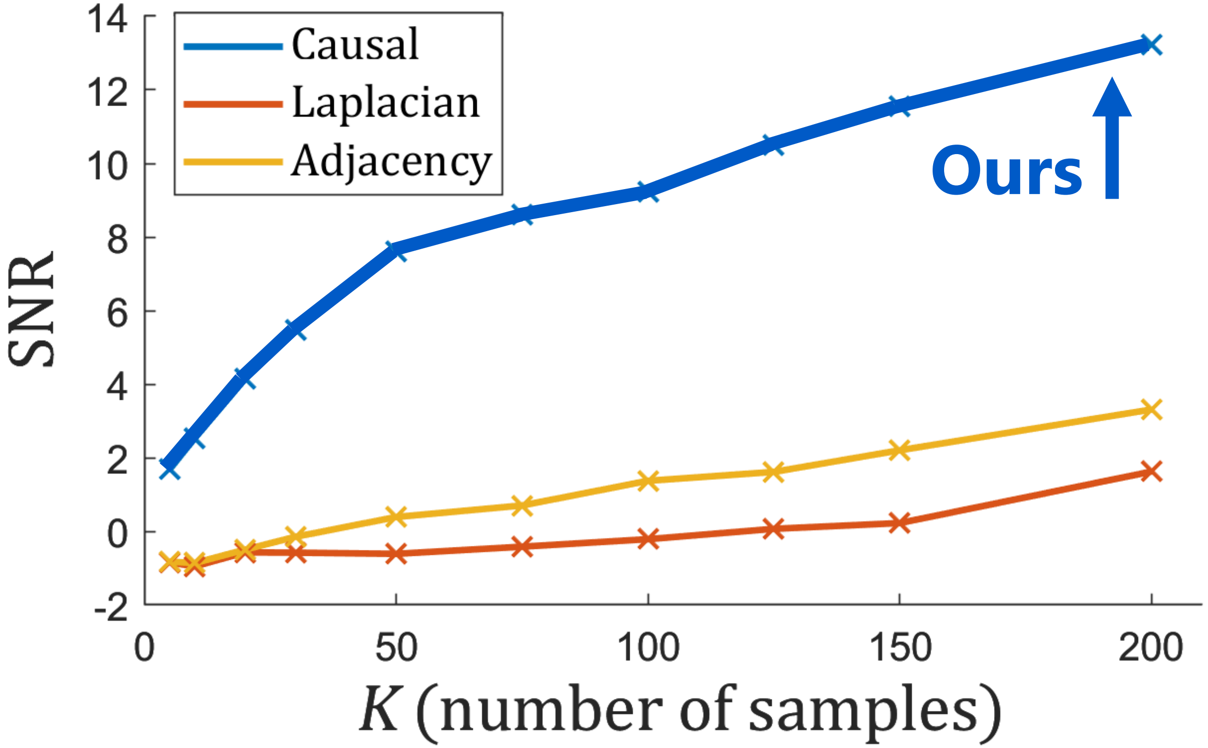
Reconstruction with other Fourier bases

Signal to noise ratio (**SNR**) as metric

Results

Best reconstruction using **our** basis

Signal is not sparse w.r.t. other bases



Problem 2: Learning the graph from data

Settings

Similar as Problem 1

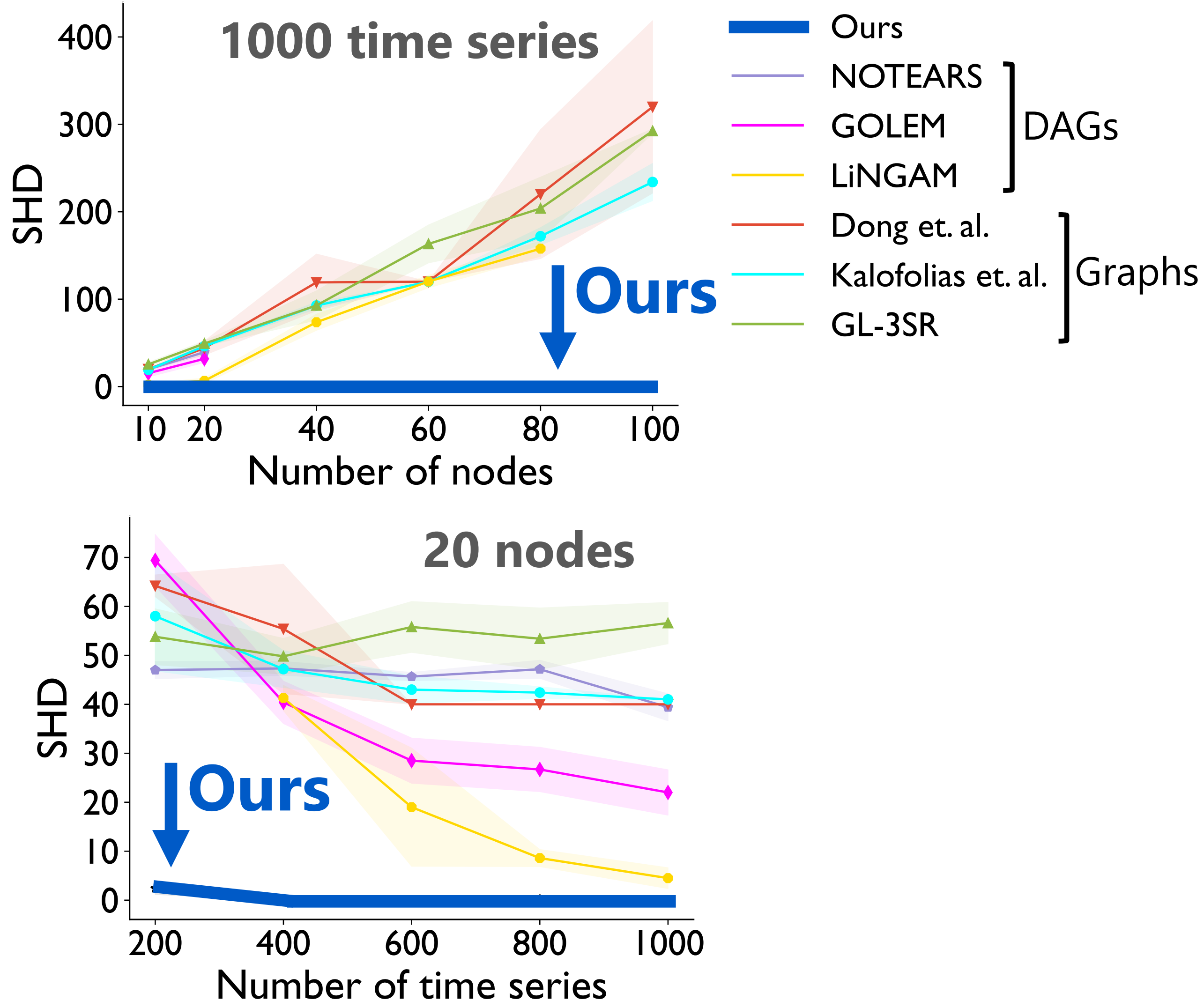
Comparison

Graph learning and DAG learning methods

SHD metric (Structural Hamming distance) lower is better

Results

Our method fits best the data generation (few causes) assumption.



A real world scenario: Thames river dataset

Settings

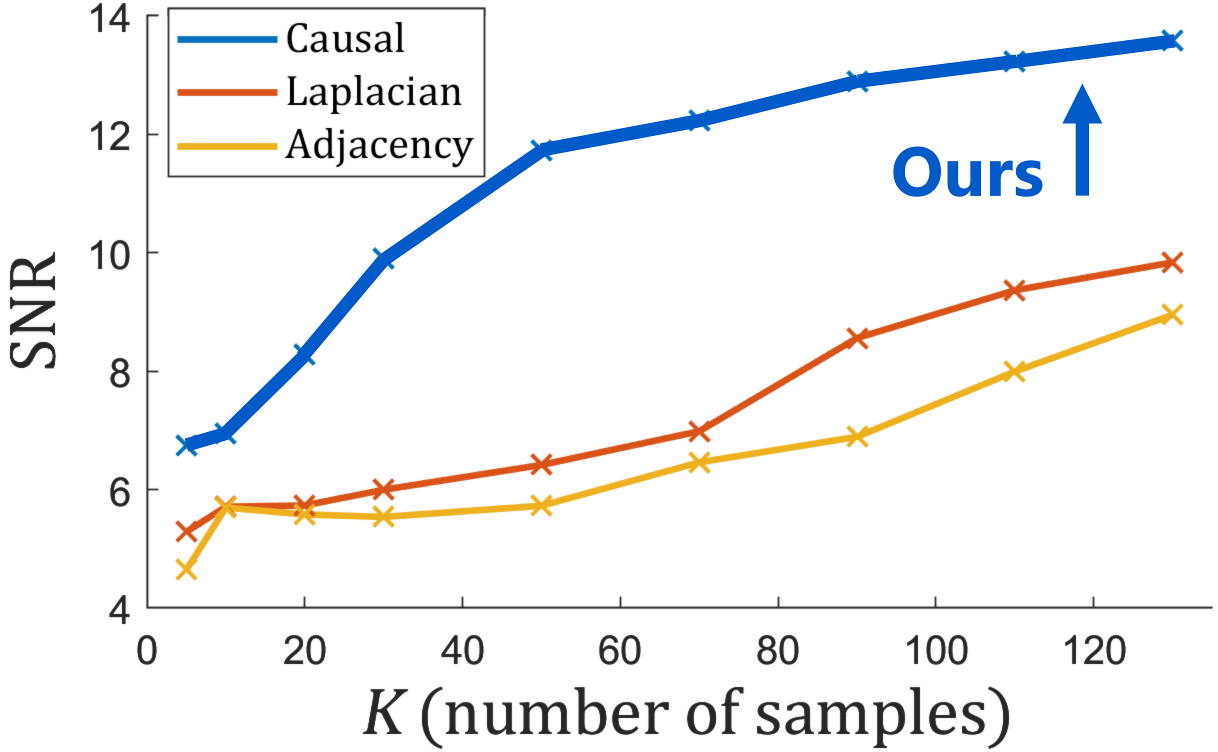
Graph: river network (unweighted), 13 nodes
Data: 7 time series, 49 time steps

Prototypical application
(Graph & signal learning)

1. Compute weights (uses all data)
2. Signal reconstruction (sampled data) using weighted causal basis

Results

Best reconstruction (compression) using **causal basis**



1. Time-series graph data

$\mathbf{X}$

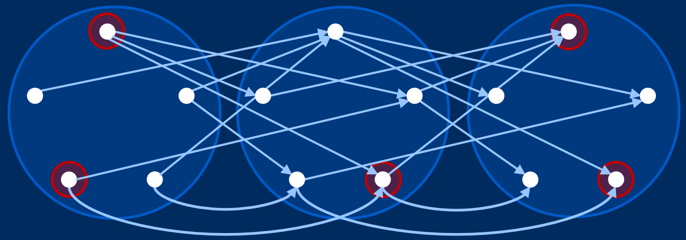
Date	INTC	MSFT	NVDA	GOOG	AAPL
2023-01-03	26.7	239.6	143.1	89.7	125.1
2023-01-04	27.7	229.1	147.5	88.7	126.4
2023-01-05	27.6	222.3	142.6	86.8	125.0
2023-01-06	28.7	224.9	148.6	88.2	129.6
2023-01-09	29.3	227.1	156.3	88.8	130.1
⋮	⋮				

B

Time = causal order

2. Unrolled into DAG

A



Impose Few causes C assumption

3. Few-cause signal reconstruction

$$\hat{\mathbf{x}} = \mathbf{W} \cdot \arg\min_{\mathbf{c}} \frac{1}{2K} \|\mathbf{\Phi}_K \mathbf{x} - \mathbf{\Phi}_K \mathbf{W}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1$$

4. Import DAG learning methods from ML to do graph learning

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}\in\mathbb{R}^{dT\times dT}} \|\mathbf{X}-\mathbf{XA}\|_1 + \lambda\|\mathbf{A}-\mathbf{PAQ}\|_2$$

$$\mathrm{A} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \mathrm{B} & 0 & \dots & 0 & 0 \\ \vdots & \mathrm{B} & \ddots & \vdots & \\ & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \dots & \mathrm{B} & 0 \end{pmatrix}$$

Learn the entire A and then extract B