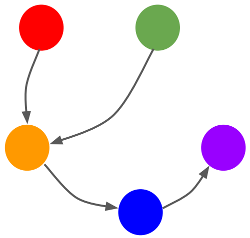
Causal Discovery



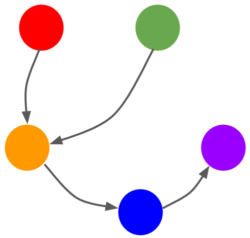
Luca Castri, PhD student @ University of Lincoln, UK

Outline

What is Causal Discovery?

Constraint-based methods
- Causal assumptions
- Markov Equivalence Class
- PC algorithm
- FCI algorithm

Causal Discovery with Time-series data
- PCMCI algorithm
- Robotics application



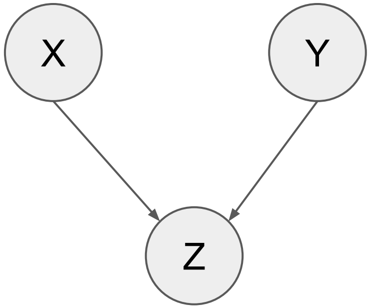
Causal Discovery

Causal Discovery: data→causal graph

	X	Y	Z
0	20.000000	100.0	340.000000
1	20.204082	100.0	340.408163
2	20.408163	100.0	340.816327
3	20.612245	100.0	341.224490
4	20.816327	100.0	341.632653
...	...	...	...
2495	29.183673	300.0	958.367347
2496	29.387755	300.0	958.775510
2497	29.591837	300.0	959.183673
2498	29.795918	300.0	959.591837
2499	30.000000	300.0	960.000000

2500 rows × 3 columns

→



Outline

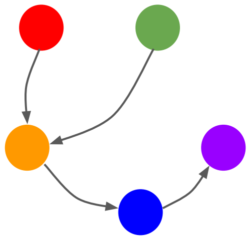
What is Causal Discovery?

Constraint-based methods

- Causal assumptions
- Markov Equivalence Class
- PC algorithm
- FCI algorithm

Causal Discovery with Time-series data

- PCMCI algorithm
- Robotics application



Constraint-based methods

They use conditional independence tests (constraints) to measure the independence between variables in order to identify the causal graph

PC - Peter Spirtes and **C**lark Glymour

FCI - Fast **C**ausal **I**nference

What are the assumptions?

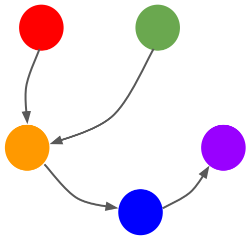
- **Markov**
- **Faithfulness**
- **Causal Sufficiency**
- **Acyclicity**

Outline

What is Causal Discovery?

Constraint-based methods
- **Causal assumptions**
- Markov Equivalence Class
- PC algorithm
- FCI algorithm

Causal Discovery with Time-series data
- PCMCI algorithm
- Robotics application



d-separation

Two (sets of) nodes X and Y are d-separated by a set of nodes Z if all the path between (any node in) X and (any node in) Y are blocked by Z

$$X \perp\!\!\!\perp_G Y | Z \quad \Rightarrow \quad X \perp\!\!\!\perp_P Y | Z$$

Markov assumption

d-separation in the graph implies conditional independencies in the distribution. **Causal graph ⇒ data**

Constraint-based methods

Causal Assumptions

d-separation

Two (sets of) nodes X and Y are d-separated by a set of nodes Z if all the path between (any node in) X and (any node in) Y are blocked by Z

$$X \perp\!\!\!\perp_G Y | Z \quad \Rightarrow \quad X \perp\!\!\!\perp_P Y | Z$$

Markov assumption

d-separation in the graph implies conditional independencies in the distribution. **Causal graph ⇒ data**

$$X \perp\!\!\!\perp_G Y | Z \quad \Leftarrow \quad X \perp\!\!\!\perp_P Y | Z$$

Faithfulness

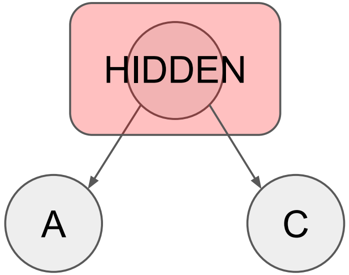
To do causal discovery we need to go the other way around.
Causal graph ⇐ data

It is the inverse of the Markov assumption and assesses:

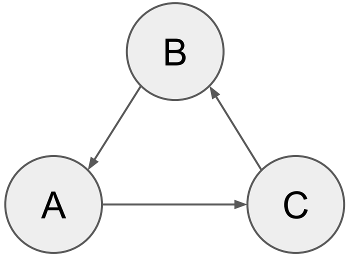
if two variables X and Y are independent conditioning on Z in the distribution then X and Y are d-separated by Z in the causal graph

Constraint-based methods
Causal Assumptions

Causal Sufficiency
There are no unobserved confounders of any of the variables in the graph



Acyclicity
No cycle in the graph



Outline

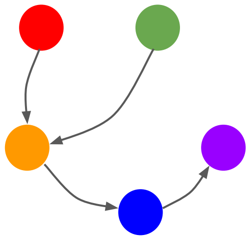
What is Causal Discovery?

Constraint-based methods

- Causal assumptions
- **Markov Equivalence Class**
- PC algorithm
- FCI algorithm

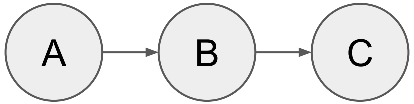
Causal Discovery with Time-series data

- PCMCI algorithm
- Robotics application

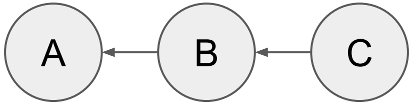


Constraint-based methods
Markov Equivalence Class

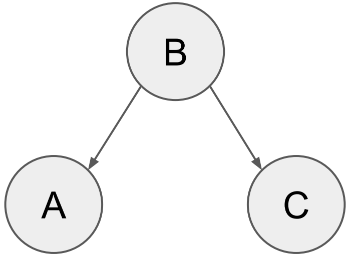
Chain



Chain



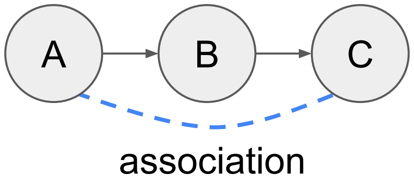
Fork



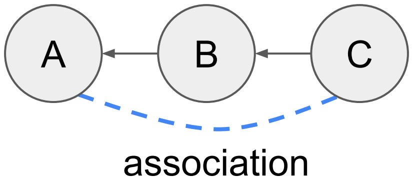
Constraint-based methods

Markov Equivalence Class

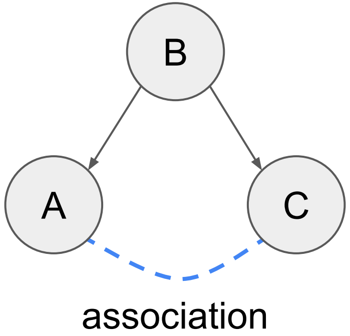
Chain



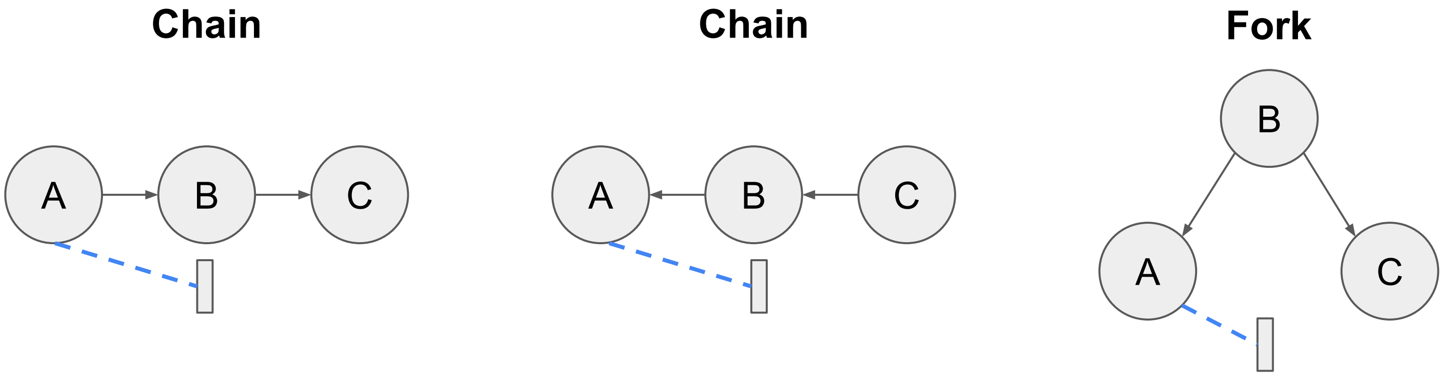
Chain



Fork



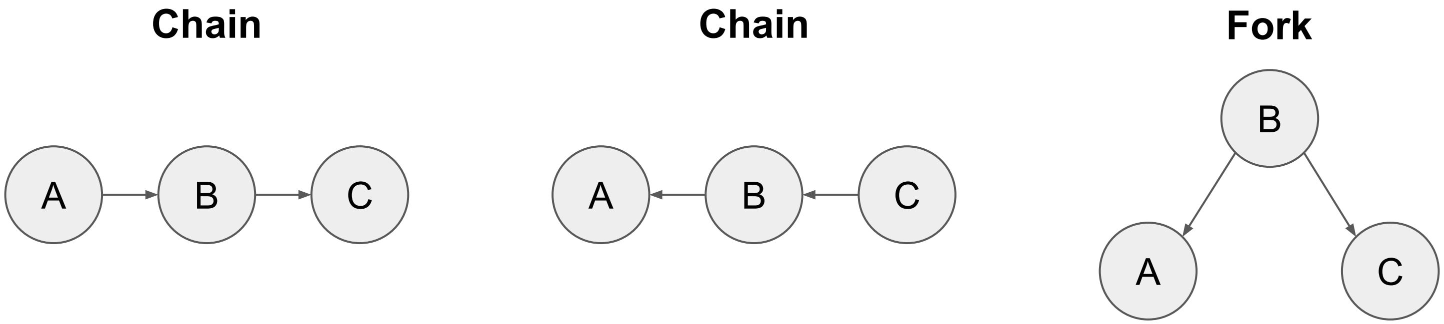
Constraint-based methods
Markov Equivalence Class



Markov: $A \perp\!\!\!\perp C | B$ They all imply the same conditional independence

Constraint-based methods

Markov Equivalence Class



Markov equivalent
They all belong to the same **Markov Equivalence Class**

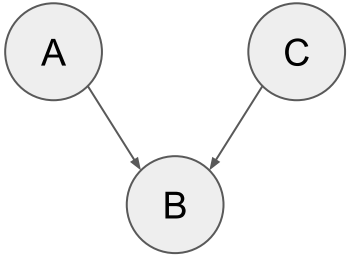
Markov: $A \perp\!\!\!\perp C | B$ They all imply the same conditional independence

Constraint-based methods

Markov Equivalence Class

Markov: $A \perp\!\!\!\perp C$

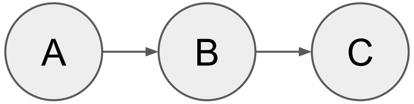
Collider

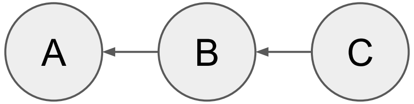


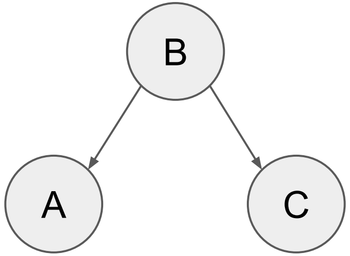
Constraint-based methods

Markov Equivalence Class

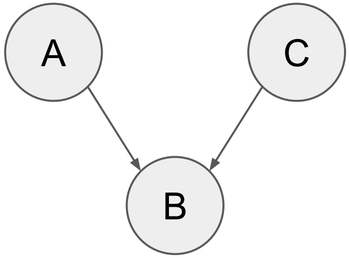
Markov: $A \perp\!\!\!\perp C \mid B$







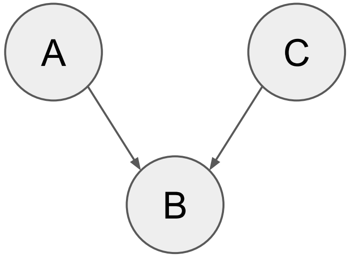
Markov: $A \perp\!\!\!\perp C$



Constraint-based methods

Markov Equivalence Class

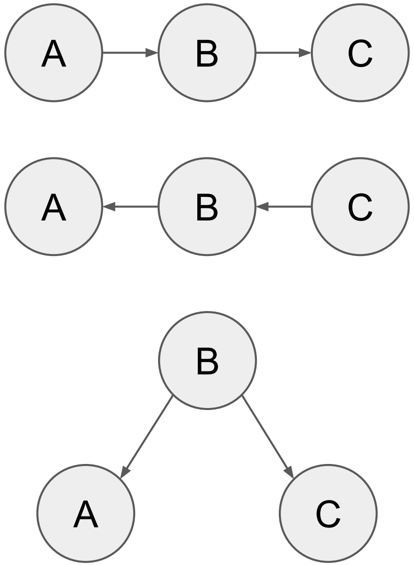
Markov: $A \perp\!\!\!\perp C$



If we find a triple of variables which satisfies this Markov assumption
$\Rightarrow$ collider configuration

COLLIDERS ARE IMPORTANT

Markov: $A \perp\!\!\!\perp C | B$



Outline

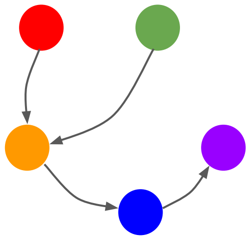
What is Causal Discovery?

Constraint-based methods

- Causal assumptions
- Markov Equivalence Class
- **PC algorithm**
- FCI algorithm

Causal Discovery with Time-series data

- PCMCI algorithm
- Robotics application



Constraint-based methods

PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

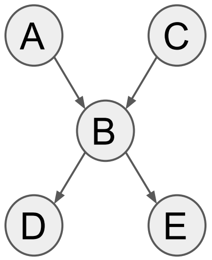
1. Start with fully connected and undirected graph
2. Identify the skeleton
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)

Constraint-based methods

PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. Identify the skeleton
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)

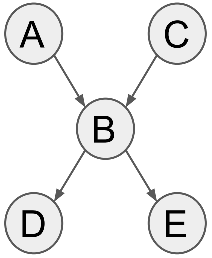


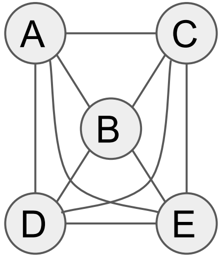
Constraint-based methods

PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. **Start with fully connected and undirected graph**
2. Identify the skeleton
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)



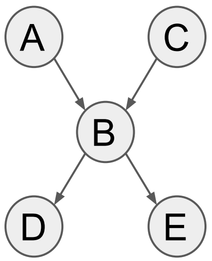


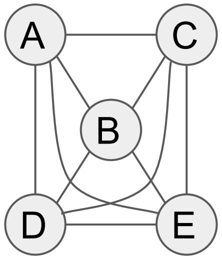
Constraint-based methods

PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. **Identify the skeleton**
   remove edges X – Y where $X \perp\!\!\!\perp Y | Z$ starting with Z empty and increasing its size
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)



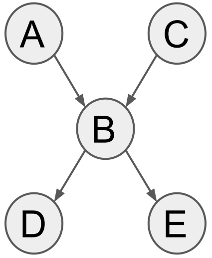


Constraint-based methods

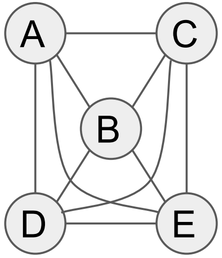
PC algorithm

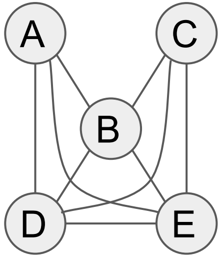
Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. **Identify the skeleton**
   remove edges X – Y where $X \perp\!\!\!\perp Y | Z$ starting with Z empty and increasing its size
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)



$A \perp\!\!\!\perp C | \{\}$

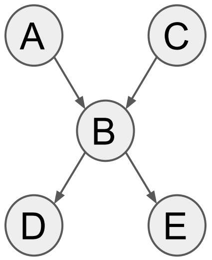




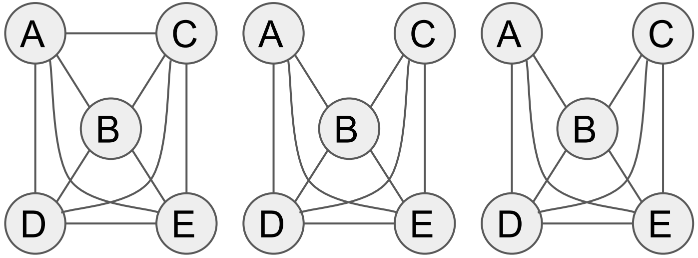
Constraint-based methods
PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. **Identify the skeleton**
   remove edges X – Y where $X \perp\!\!\!\perp Y | Z$ starting with Z empty and increasing its size
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)



$A \perp\!\!\!\perp C | \{\}$



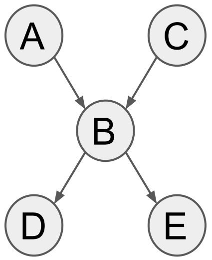
$A \perp\!\!\!\perp D | \{\}$ ✖
$D \perp\!\!\!\perp E | \{\}$ ✖
$C \perp\!\!\!\perp E | \{\}$ ✖
$A \perp\!\!\!\perp E | \{\}$ ✖
$C \perp\!\!\!\perp D | \{\}$ ✖

Constraint-based methods

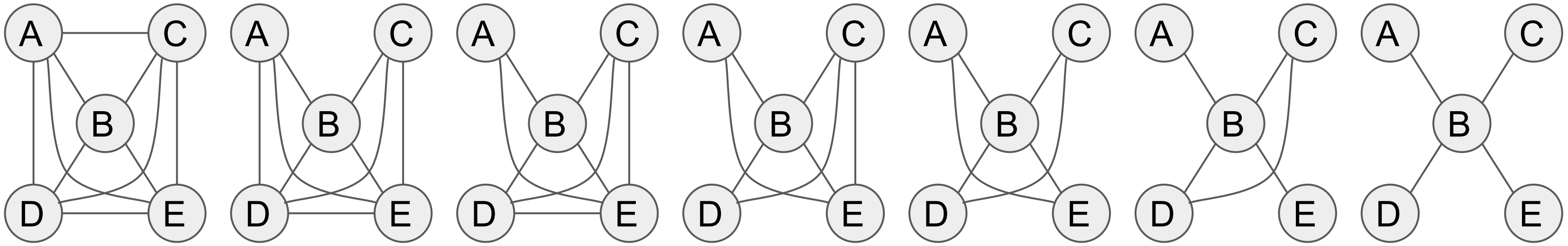
PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. **Identify the skeleton**
   remove edges X – Y where $X \perp\!\!\!\perp Y | Z$ starting with Z empty and increasing its size
3. Identify colliders and orient them
4. Orient the non-colliders edges (orientation propagation)



$A \perp\!\!\!\perp C|\{\}$ $\quad A \perp\!\!\!\perp D|\{B\}$ $\quad D \perp\!\!\!\perp E|\{B\}$ $\quad C \perp\!\!\!\perp E|\{B\}$ $\quad A \perp\!\!\!\perp E|\{B\}$ $\quad C \perp\!\!\!\perp D|\{B\}$

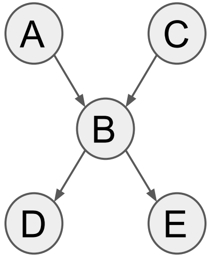


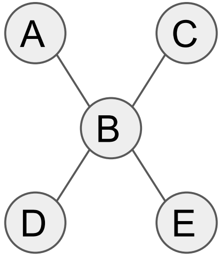
Constraint-based methods

PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. Identify the skeleton
3. **Identify colliders and orient them**
   for any path X – Z – Y where there is no edge between
   X and Y and, Z was never included in the conditioning set
   ⇒ X→Z←Y collider
4. Orient the non-colliders edges (orientation propagation)



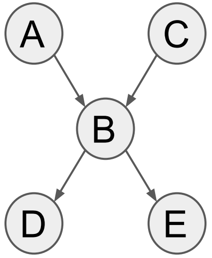


Constraint-based methods

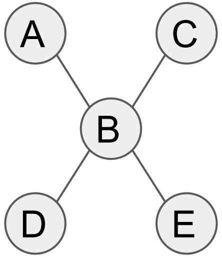
PC algorithm

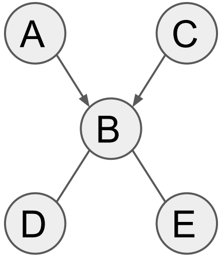
Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. Identify the skeleton
3. **Identify colliders and orient them**
   for any path X – Z – Y where there is no edge between X and Y and, Z was never included in the conditioning set
   ⇒ X→Z←Y collider
4. Orient the non-colliders edges (orientation propagation)



$A \perp\!\!\!\perp C | \{\}$ We removed A – C with conditioning set empty
⇒ unique markov equivalence class: **collider**
⇒ A→B←C



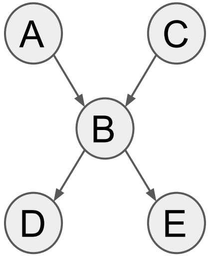


Constraint-based methods

PC algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

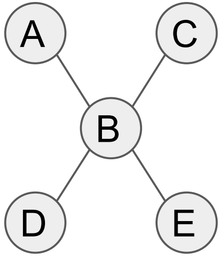
1. Start with fully connected and undirected graph
2. Identify the skeleton
3. **Identify colliders and orient them**
   for any path X – Z – Y where there is no edge between
   X and Y and, Z was never included in the conditioning set
   ⇒ X→Z←Y collider
4. Orient the non-colliders edges (orientation propagation)

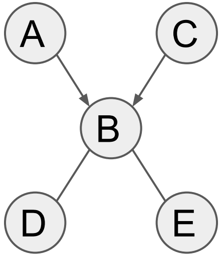


$A \perp\!\!\!\perp C | \{\}$ We removed A – C with conditioning set empty
⇒ unique markov equivalence class: **collider**
⇒ A→B←C

$A \perp\!\!\!\perp D | \{B\}$ $A \perp\!\!\!\perp E | \{B\}$ $D \perp\!\!\!\perp E | \{B\}$ $C \perp\!\!\!\perp D | \{B\}$ $C \perp\!\!\!\perp E | \{B\}$

We removed all these edges conditioning on B
⇒ markov equivalence class: **chain** or **fork**

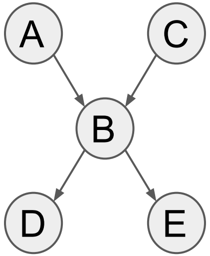


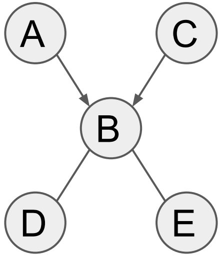


Constraint-based methods
PC algorithm

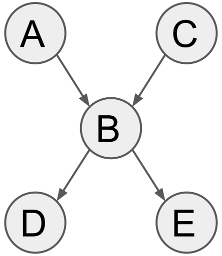
Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. Identify the skeleton
3. Identify colliders and orient them
4. **Orient the non-colliders edges (orientation propagation)**
   any edge $Z – Y$ part of a partially directed path $X \rightarrow Z – Y$, where there is no edge between X and Y can be oriented as $Z \rightarrow Y$





$A \rightarrow B – D$, no edge between A and D $\Rightarrow A \rightarrow B \rightarrow D$
$A \rightarrow B – E$, no edge between A and E $\Rightarrow A \rightarrow B \rightarrow E$

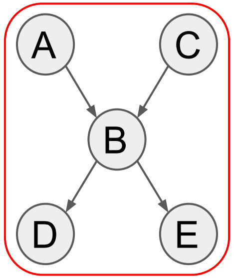


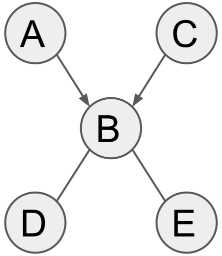
Constraint-based methods

PC algorithm

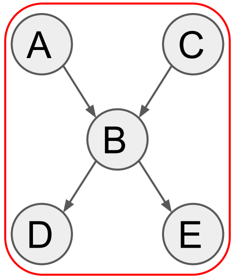
Markov | Faithfulness | Causal Sufficiency | Acyclicity

1. Start with fully connected and undirected graph
2. Identify the skeleton
3. Identify colliders and orient them
4. **Orient the non-colliders edges (orientation propagation)**
   any edge Z – Y part of a partially directed path X→Z – Y, where there is no edge between X and Y can be oriented as Z→Y





A→B – D, no edge between A and D ⇒ A→B→D
A→B – E, no edge between A and E ⇒ A→B→E



Outline

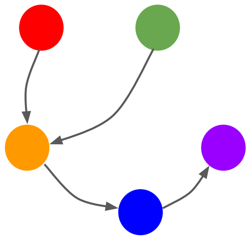
What is Causal Discovery?

Constraint-based methods

- Causal assumptions
- Markov Equivalence Class
- PC algorithm
- **FCI algorithm**

Causal Discovery with Time-series data

- PCMCI algorithm
- Robotics application

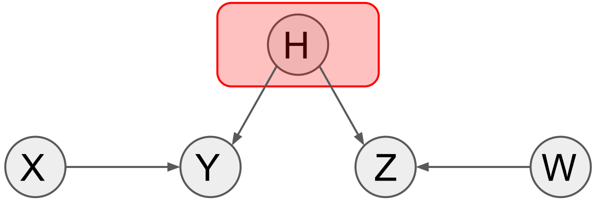


Constraint-based methods

FCI algorithm

Markov | Faithfulness | Causal Sufficiency | Acyclicity

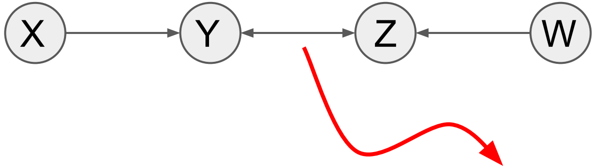
variation of the PC algorithm which tolerates hidden confounders.



skeleton

colliders identification

colliders identification



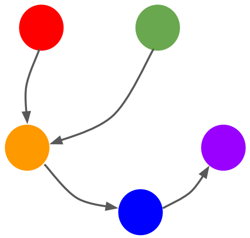
The bidirected edge indicates that there is at least one <u>hidden confounders</u>

Outline

What is Causal Discovery?

Constraint-based methods
- Causal assumptions
- Markov Equivalence Class
- PC algorithm
- FCI algorithm

Causal Discovery with Time-series data
- PCMCI algorithm
- Robotics application

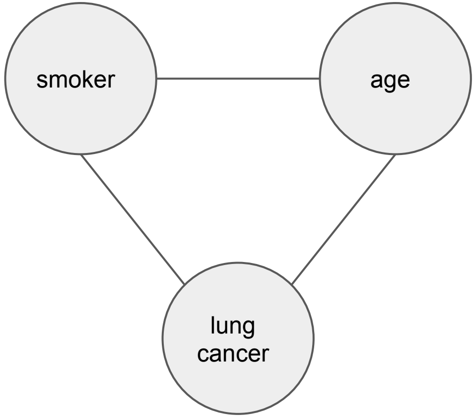


Causal Discovery with Time-series data

The PC and FCI causal discovery method work well with discrete/categorical data.

example
non-smoker **0** - smoker **1**
age under 50 **0** - age over 50 **1**
no lung cancer **0** - lung cancer **1**

Smoker	Age	Lung cancer
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1



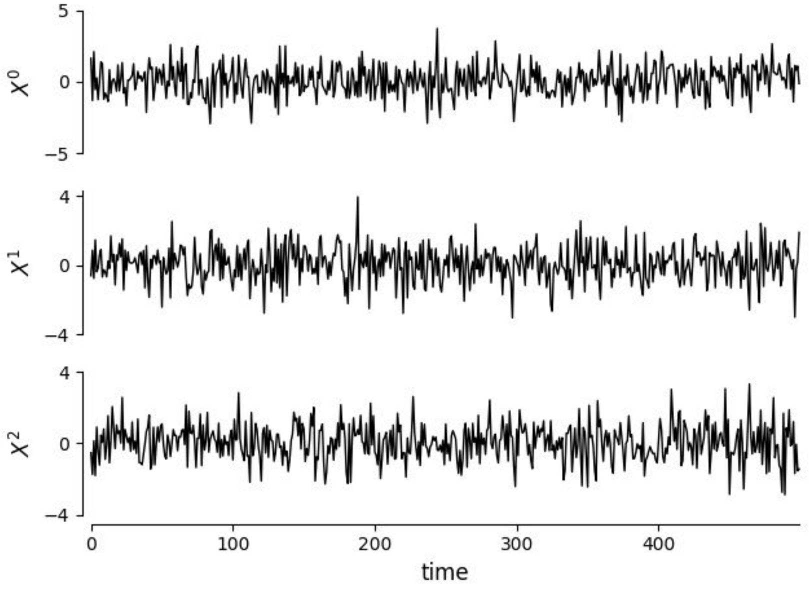
Causal Discovery with Time-series data

The PC and FCI causal discovery method work well with discrete/categorical data.

What if we deal with time-series data?
PC is inappropriate to use with time series data due to:

- time ordering
- lagged dependencies
- high false positive rates due to the autocorrelation.

$$\begin{cases} X_t^0 = 0.2(X_{t-1}^1)^2 + \eta_t^0 \\ X_t^1 = \eta_t^1 \\ X_t^2 = 0.3(X_{t-2}^1)^2 + \eta_t^2 \end{cases}$$

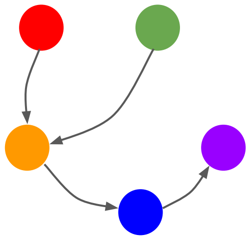


Outline

What is Causal Discovery?

Constraint-based methods
- Causal assumptions
- Markov Equivalence Class
- PC algorithm
- FCI algorithm

Causal Discovery with Time-series data
- **PCMCI algorithm**
- Robotics application



Causal Discovery with Time-series data

PCMCI algorithm

It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t | \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay

Causal Discovery with Time-series data
PCMCI algorithm

It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t | \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay

Causal Discovery with Time-series data

PCMCI algorithm

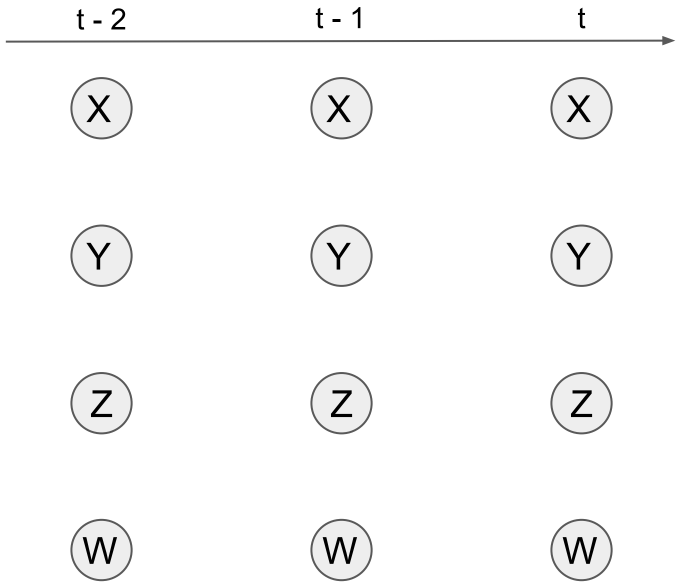
It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t \,|\, \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay



Causal Discovery with Time-series data

PCMCI algorithm

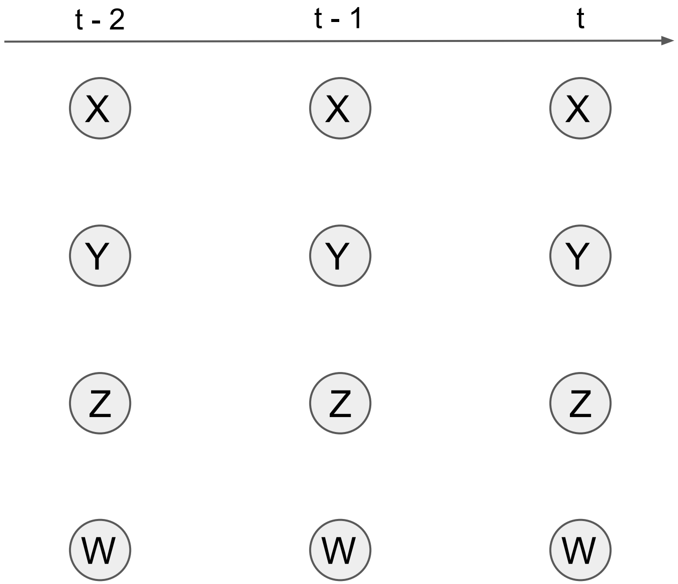
It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t | \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay



Causal Discovery with Time-series data

PCMCI algorithm

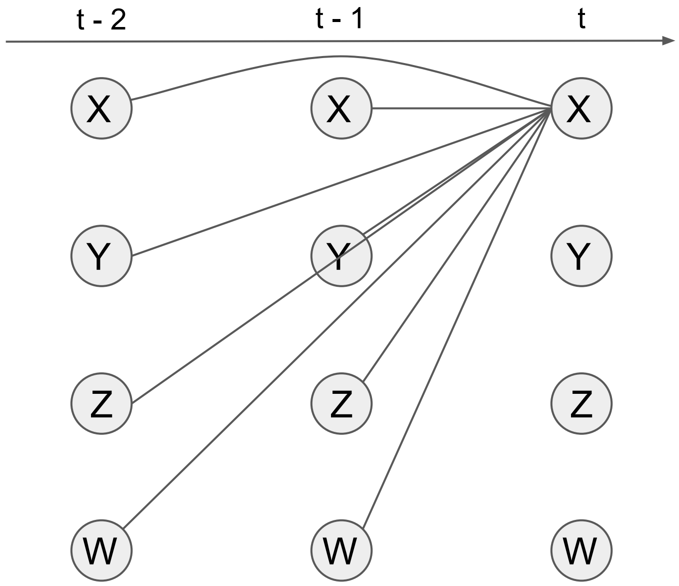
It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t | \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay



Causal Discovery with Time-series data

PCMCI algorithm

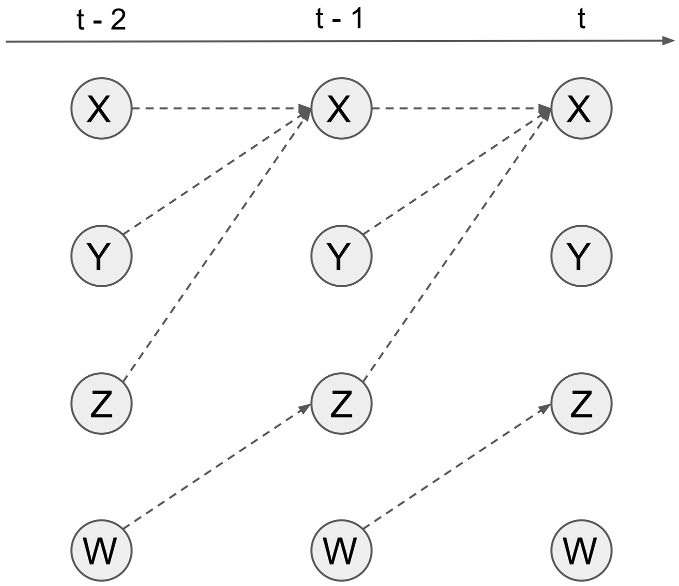
It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t | \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay



Causal Discovery with Time-series data

PCMCI algorithm

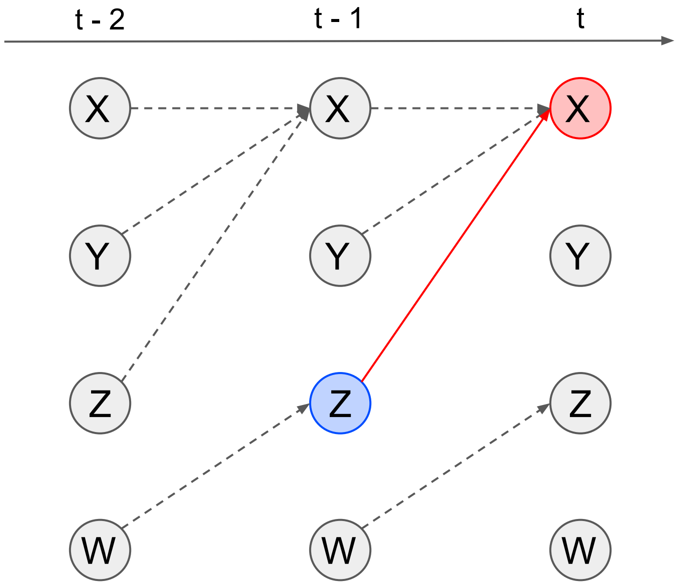
It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$\boxed{X^i_{t-\tau}} \perp\!\!\!\perp \boxed{X^j_t} | \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay



Causal Discovery with Time-series data

PCMCI algorithm

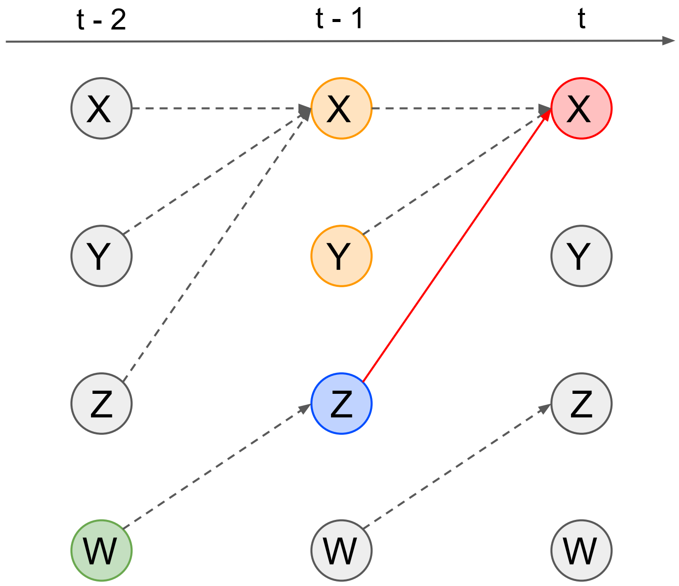
It consists of two main steps:

- **PC algorithm**
  retrieves the causal model structure by considering ONLY lagged dependencies as possible causal relationships between variables

- **MCI test**
  validates the structure found at the previous step by performing a false positive rate optimisation control

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t \mid \tilde{P}(X^i_{t-\tau}), \tilde{P}(X^j_t)$$

Key parameter: τ maximum time delay



Causal Discovery with Time-series data

PCMCI algorithm

$$\begin{cases} X_t^0 = 0.4(X_{t-1}^1)^2 + \eta_t^0 \\ X_t^1 = \eta_t^1 \\ X_t^2 = 0.3(X_{t-2}^1)^2 + \eta_t^2 \end{cases}$$

```
random_state = np.random.default_rng(seed=42)
data = random_state.standard_normal((500, 3))
for t in range(1, 500):
    data[t, 0] += 0.4*data[t-1, 1]**2
    data[t, 2] += 0.3*data[t-2, 1]**2
var_names = [r'$X^0$', r'$X^1$', r'$X^2$']

dataframe = pp.DataFrame(data, var_names=var_names)
```

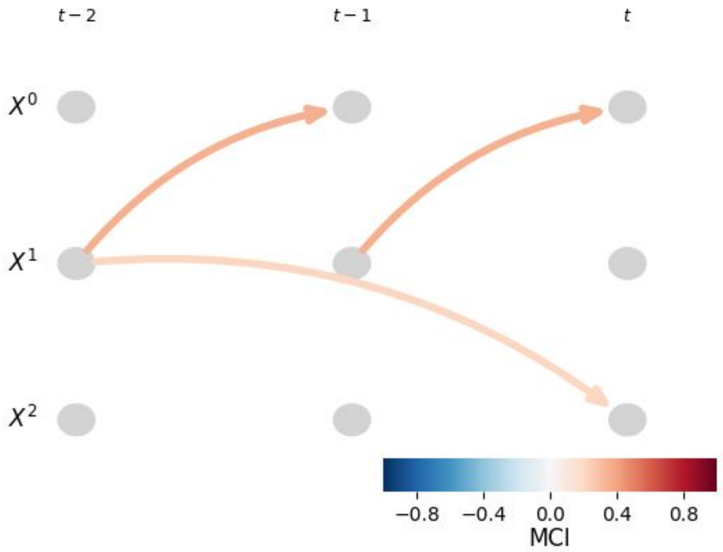
```
gpdc = GPDC(significance='analytic', gp_params=None)
pcmci_gpdc = PCMCI(
    dataframe=dataframe,
    cond_ind_test=gpdc,
    verbosity=0)
```

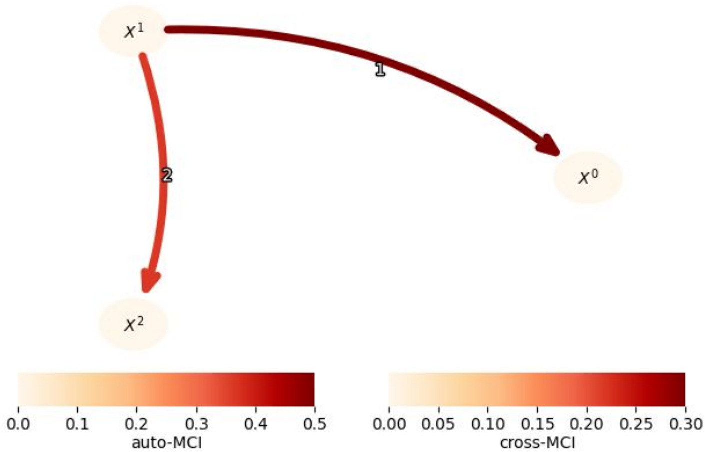
```
results = pcmci_gpdc.run_pcmci(tau_max=2, pc_alpha=0.1, alpha_level = 0.01)
tp.plot_graph(
    val_matrix=results['val_matrix'],
    graph=results['graph'],
    var_names=var_names,
    show_colorbar=False,
    )
```

Causal Discovery with Time-series data

PCMCI algorithm

$$
\begin{cases}
X_t^0 = 0.4(X_{t-1}^1)^2 + \eta_t^0 \\
X_t^1 = \eta_t^1 \\
X_t^2 = 0.3(X_{t-2}^1)^2 + \eta_t^2
\end{cases}
$$





Outline

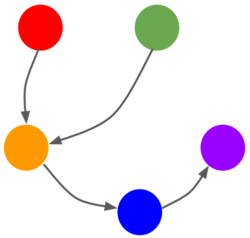
What is Causal Discovery?

Constraint-based methods

- Causal assumptions
- Markov Equivalence Class
- PC algorithm
- FCI algorithm

Causal Discovery with Time-series data

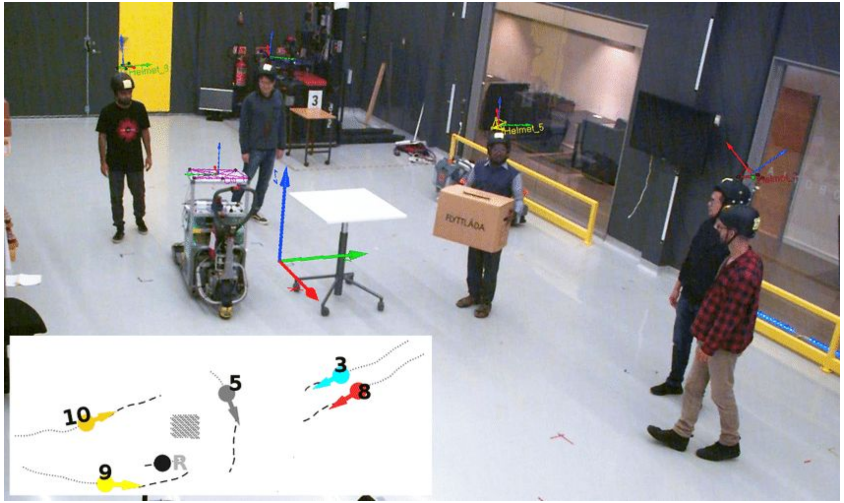
- PCMCI algorithm
- **Robotics application**

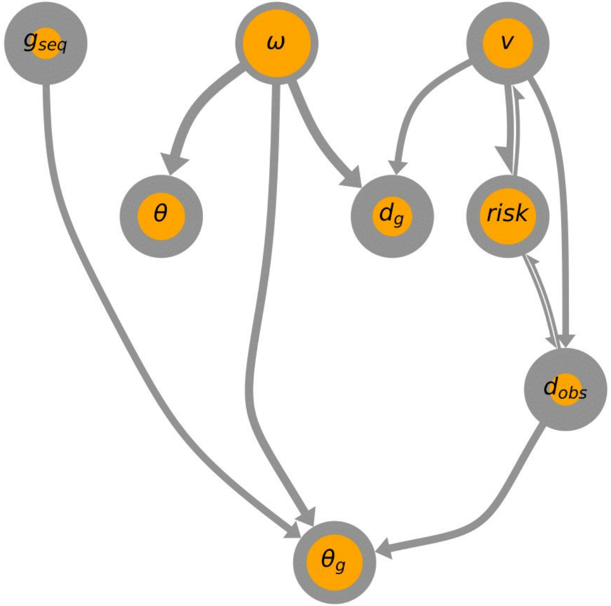


Causal Discovery with Time-series data

Robotics application

Only observational data



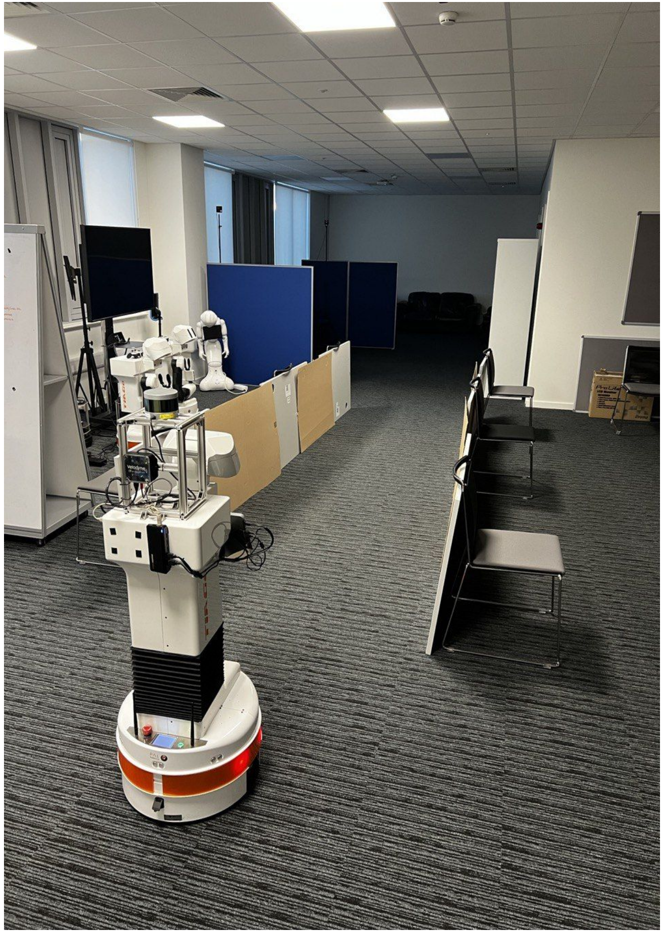


Castri, Luca, et al. "Causal Discovery of Dynamic Models for Predicting Human Spatial Interactions." (2022).
Castri, Luca, et al. "Enhancing Causal Discovery from Robot Sensor Data in Dynamic Scenarios." (2023).

Causal Discovery with Time-series data
Robotics application

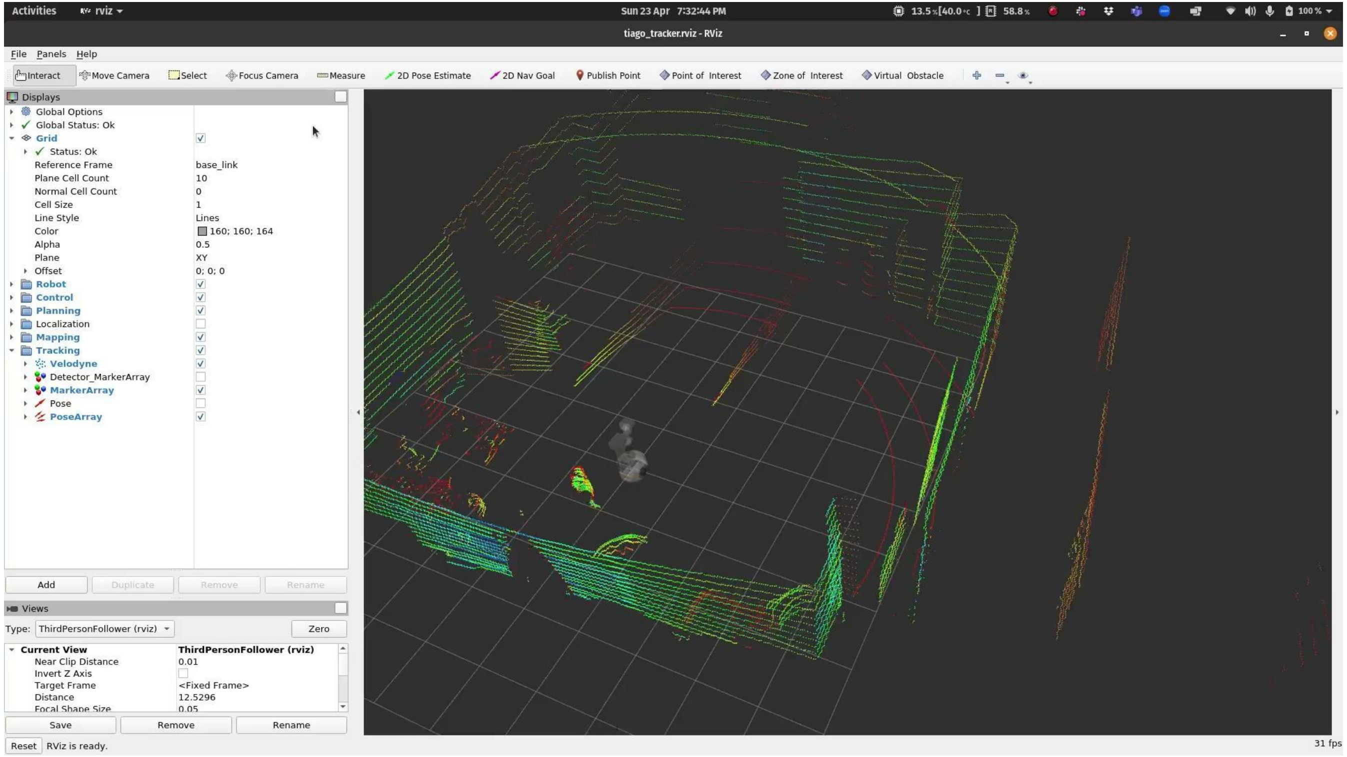
Observational and interventional data
The experiment consists of:

- TIAGo robot
- a human
- a “corridor”
- human overtakes the robot
  the robot tries to perform an action (intervention) to facilitate the overtake



Causal Discovery with Time-series data

Robotics application



CAUSAL MODEL

?

still ongoing..

Reference

- Pearl, J. 2009, Causality: Models, Reasoning and Inference, Cambridge University Press.
- Glymour, C., Zhang, K. and Spirtes, P., 2019. Review of causal discovery methods based on graphical models. Frontiers in genetics, 10, p.524.
- Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D., 2019. Detecting and quantifying causal associations in large nonlinear time series datasets. Science advances, 5(11).
- Castri, L., Mghames, S., Hanheide, M. and Bellotto, N., 2023. Enhancing Causal Discovery from Robot Sensor Data in Dynamic Scenarios. In 2nd Conference on Causal Learning and Reasoning.
- Castri, L., Mghames, S., Hanheide, M. and Bellotto, N., 2023. Causal discovery of dynamic models for predicting human spatial interactions. In Social Robotics: 14th International Conference, ICSR 2022, Italy.

Thank you questions?