Causal Discovery from Interventions

Luca Bernardinello

Università degli studi di Milano–Bicocca
Dottorato di ricerca in informatica
CAUSAL NETWORKS

24 settembre 2021

Causal discovery from interventions

Problem

Discover (identify) the “true” causal graph of a given phenomenon on a given set of variables with observational data

Assumptions

- Causal Markov condition: every vertex X in the graph is independent of its non-descendants, given its parents
- Faithfulness: if X is independent from Y given C in the probability distribution, then X is d-separated from Y given C in the causal graph
- Acyclicity
- Causal sufficiency: no unmeasured causes of any pair of variables

Causal discovery from interventions

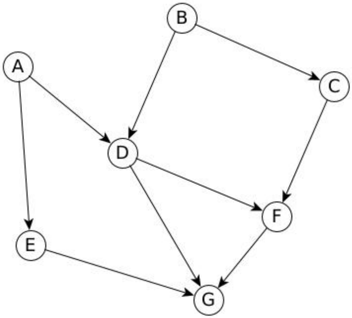
Problem

Discover (identify) the "true" causal graph of a given phenomenon on a given set of variables with observational data

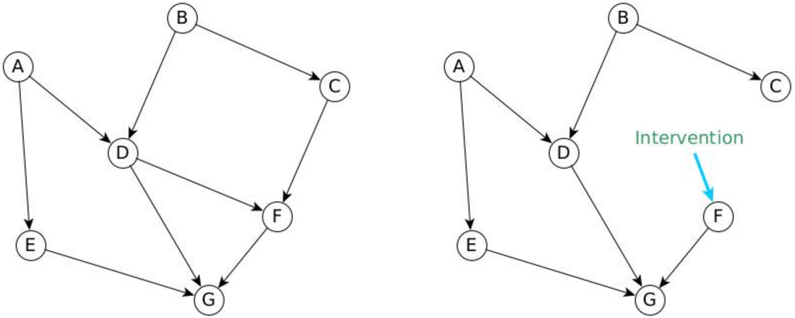
Result: a DAG (or a PDAG)

Technique: apply interventions on variables; an important difference from last lecture: interventions change the causal graph

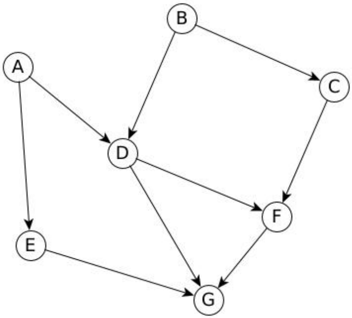
Interventions



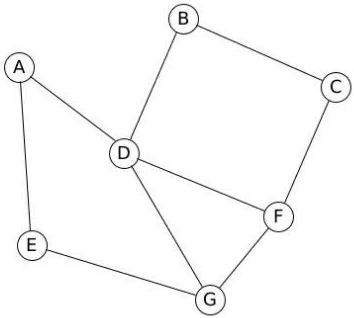
Interventions



Essential graph



Skeleton



Immoralities: $[A, B, D]$, $[C, D, F]$, $[D, E, G]$, $[E, F, G]$

A simple case: two variables

How many different causal graphs on n variables?

$$f(n) = \sum_{i=1}^{n} (-1)^{i+1} \frac{n!}{(n-i)!i!} 2^{i(n-1)} f(n-i)$$

For $n = 2$, three different causal graphs

Two variable graphs

Three distinct causal graphs

1

2

3

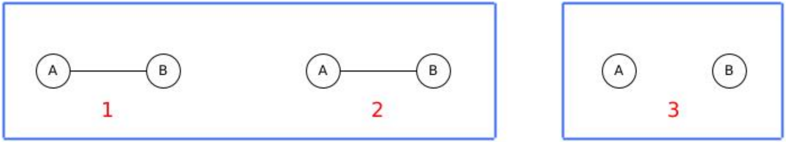
Essential graphs

1

2

3

Equivalence classes



Detectable with no intervention

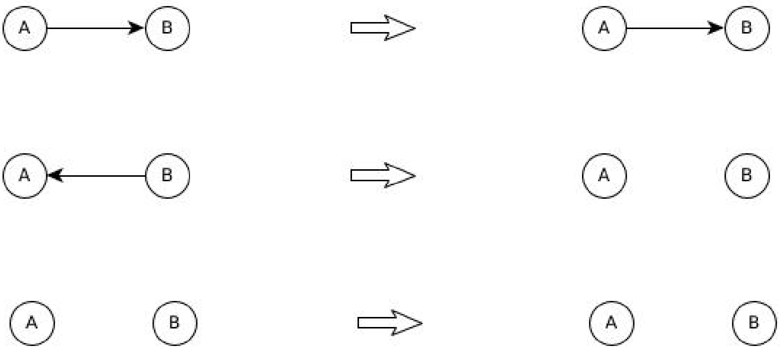
Two variables – single node intervention

What happens if we intervene on one variable?

Two variables – single node intervention

What happens if we intervene on one variable?

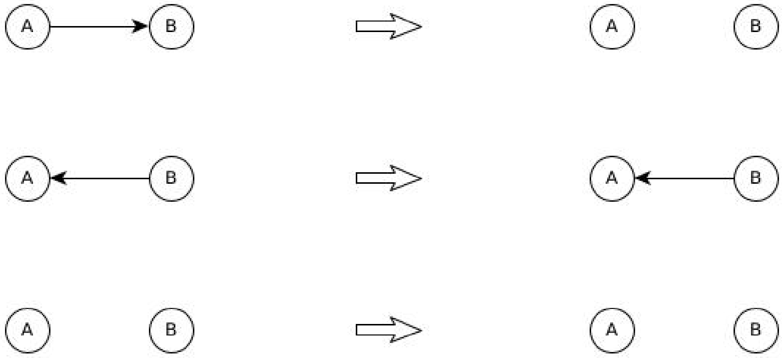
Intervention variable: A



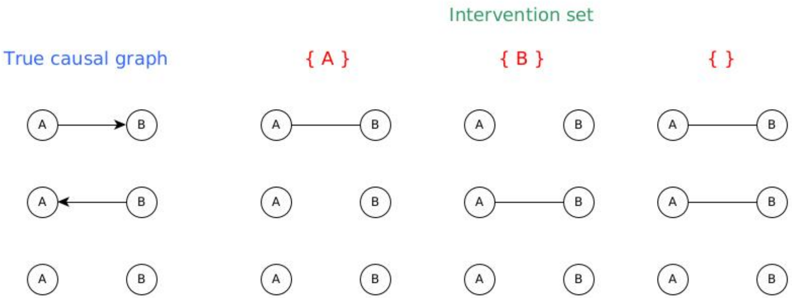
Two variables – single node intervention

What happens if we intervene on one variable?

Intervention variable: B



Two variables – single node intervention



Two variables – single node intervention

- No intervention = observational data $\Rightarrow$ skeleton

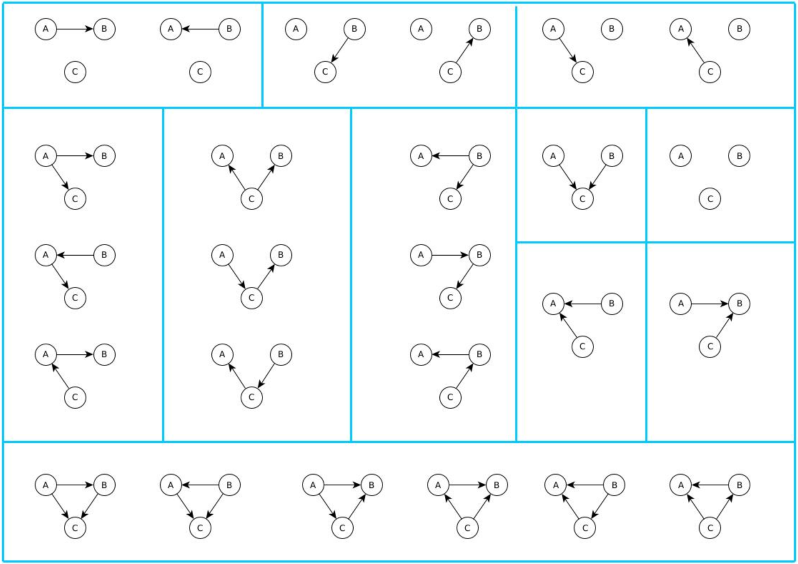
Two variables – single node intervention

- No intervention = observational data $\Rightarrow$ skeleton
- Skeleton + one intervention $\Rightarrow$ full causal graph

Two variables – single node intervention

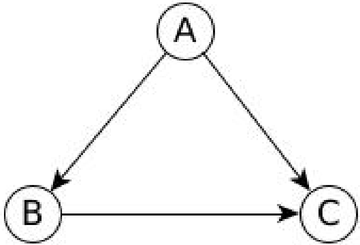
- No intervention = observational data $\Rightarrow$ skeleton
- Skeleton + one intervention $\Rightarrow$ full causal graph
- Two single node interventions $\Rightarrow$ full causal graph

Three variables



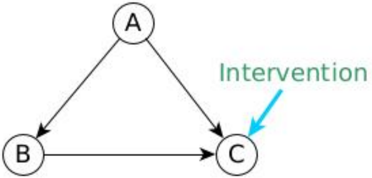
Three variables

An example with three variables



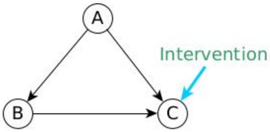
Three variables

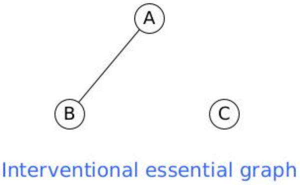
An example with three variables



Three variables

An example with three variables





Three variables

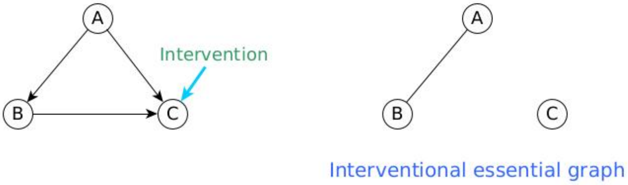
An example with three variables

Inferences:

- no arc from C to A

Three variables

An example with three variables

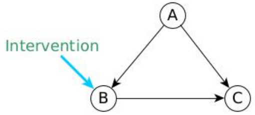


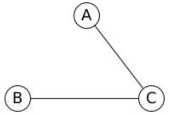
Inferences:

- no arc from C to A
- no arc from C to B

Three variables

An example with three variables

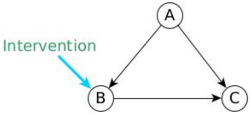


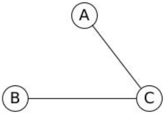


Interventional essential graph

Three variables

An example with three variables





Interventional essential graph

Inferences:

- no arc from C to B, hence arc from B to C

Three variables

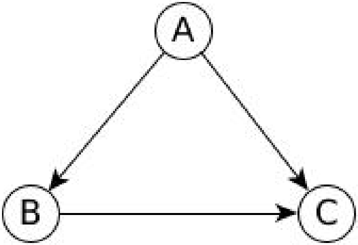
An example with three variables

Inferences:

- no arc from C to B, hence arc from B to C
- no arc from B to A

Three variables

Putting all the information together, we derive the complete causal graph:



Single node interventions – general case

Theorem Let $\mathcal{G}$ be a causal graph on $n > 2$ variables. Then $n - 1$ single node interventions are sufficient to identify $\mathcal{G}$.

Informal argument:

- The first intervention identifies the adjacencies between the other $n - 1$ nodes
- The i-th intervention directs the edges incident on X_i
  1. if $X_i \perp X_j$, then $X_i \leftarrow X_j$
  2. if $X_i \not\perp X_j$, then $X_i \rightarrow X_j$
- All edges incident on X_n have been already directed in the first $n - 1$ interventions

Complete graphs

Theorem Two graphs are Markov equivalent if and only if they have the same skeleton and the same immoralities (Verma and Pearl; Frydenburg).

Complete graphs

Theorem Two graphs are Markov equivalent if and only if they have the same skeleton and the same immoralities (Verma and Pearl; Frydenburg).

In a complete graph there is no immorality (why?), so Markov equivalence coincides with graph isomorphism. The PC algorithm will find the complete undirected graph as skeleton.

Complete graphs

Theorem Two graphs are Markov equivalent if and only if they have the same skeleton and the same immoralities (Verma and Pearl; Frydenburg).

In a complete graph there is no immorality (why?), so Markov equivalence coincides with graph isomorphism. The PC algorithm will find the complete undirected graph as skeleton.

Theorem Let $\mathcal{G}$ be a causal graph on $n > 2$ variables. Then $n - 1$ single node interventions are necessary to identify $\mathcal{G}$ in the worst case (complete graph).

Multiple node interventions

No restriction on the number of nodes per intervention. **Theorem** Let $\mathcal{G}$ be a causal graph on $n > 2$ variables. Then

- $\lfloor \log_2(n) \rfloor + 1$ multiple node interventions are sufficient to identify $\mathcal{G}$.
- $\lfloor \log_2(n) \rfloor + 1$ multiple node interventions are necessary to identify $\mathcal{G}$ in the worst case (complete graph).

Multiple node interventions

Consider now a general, non-complete, graph with n variables. Assuming no restriction on the number of nodes per intervention, and starting with the Markov equivalence class of the graph, how many interventions are necessary?

Multiple node interventions

Consider now a general, non-complete, graph with n variables. Assuming no restriction on the number of nodes per intervention, and starting with the Markov equivalence class of the graph, how many interventions are necessary?

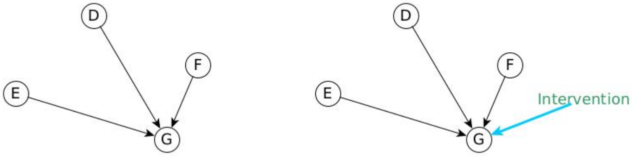
Theorem $\lceil \log_2(c) \rceil$ multi-node interventions are necessary, where c is the size of the largest clique of $\mathcal{G}$.

Parametric interventions

A *parametric intervention* (also called *soft i.*) does not change the structure of the causal graph, but modifies the conditional probabilities of a node, given its parents.

Parametric interventions

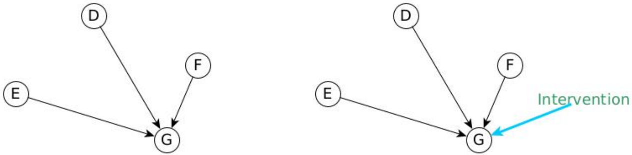
A *parametric intervention* (also called *soft i.*) does not change the structure of the causal graph, but modifies the conditional probabilities of a node, given its parents.



$$G = f_{\theta}(E, D, F) \quad \Longrightarrow \quad G = f_{\theta'}(E, D, F)$$

Parametric interventions

A *parametric intervention* (also called *soft i.*) does not change the structure of the causal graph, but modifies the conditional probabilities of a node, given its parents.



$$G = f_{\theta}(E, D, F) \quad \Longrightarrow \quad G = f_{\theta'}(E, D, F)$$

Structural (hard) interventions can be seen as a special case of parametric interventions.

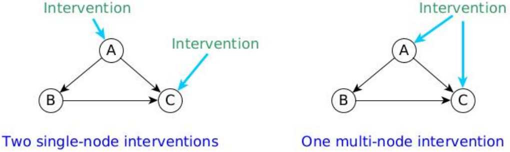
Parametric interventions

For a graph with n nodes

- $n-1$ single-node parametric interventions are sufficient for identification
- $n-1$ single-node parametric interventions are necessary for identification in the worst case

Further topics

- Partial identification with "few" interventions; how much of the causal graph can be identified with a bound number of interventions?
- Multi-node interventions



- Randomized algorithms: $O(\log \log n)$ interventions are sufficient with high probability
- Bounds on the number of variables per intervention
- Bounds on the total number of interventions
- . . .