Methods for Causal Inference
Lecture 12

Ava Khamseh
School of Informatics

2021-2022

Do Calculus

- Do-calculus: Contains, as subsets:
  - Backdoor criterion
  - Front-door criterion

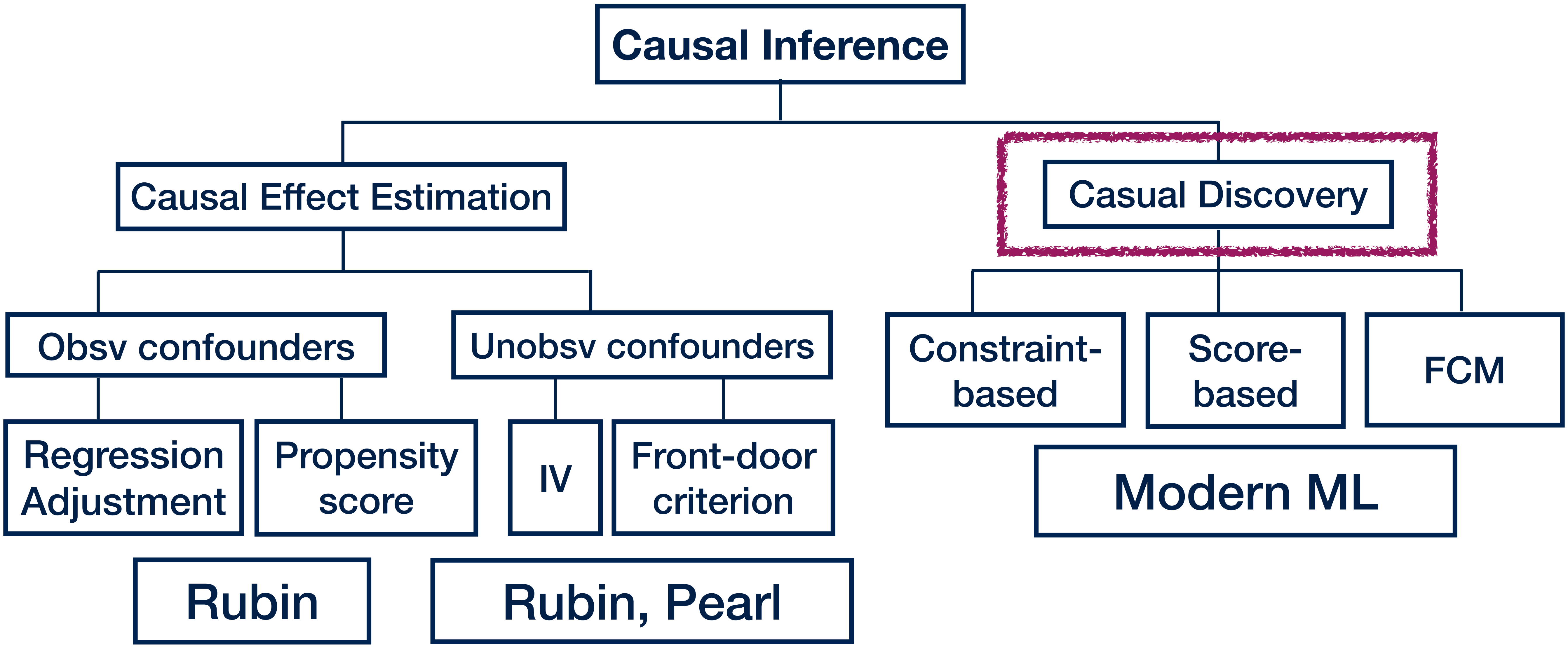
- Allows analysis of more intricate structure beyond back- and front-door

- Uncovers **all** causal effects that can be identified from a given causal graph

- Power of causal graphs is not just representation but towards **discovery** of causal information

So far …

- **Lecture 1**: Introduction & motivation, why do we care about causality?
- **Lecture 2**: Recap of probability theory, e.g., variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3**: Recap of regression, multiple regression, graphs, SCM
- **Lectures 4-20**:



Directed vs Undirected Aside

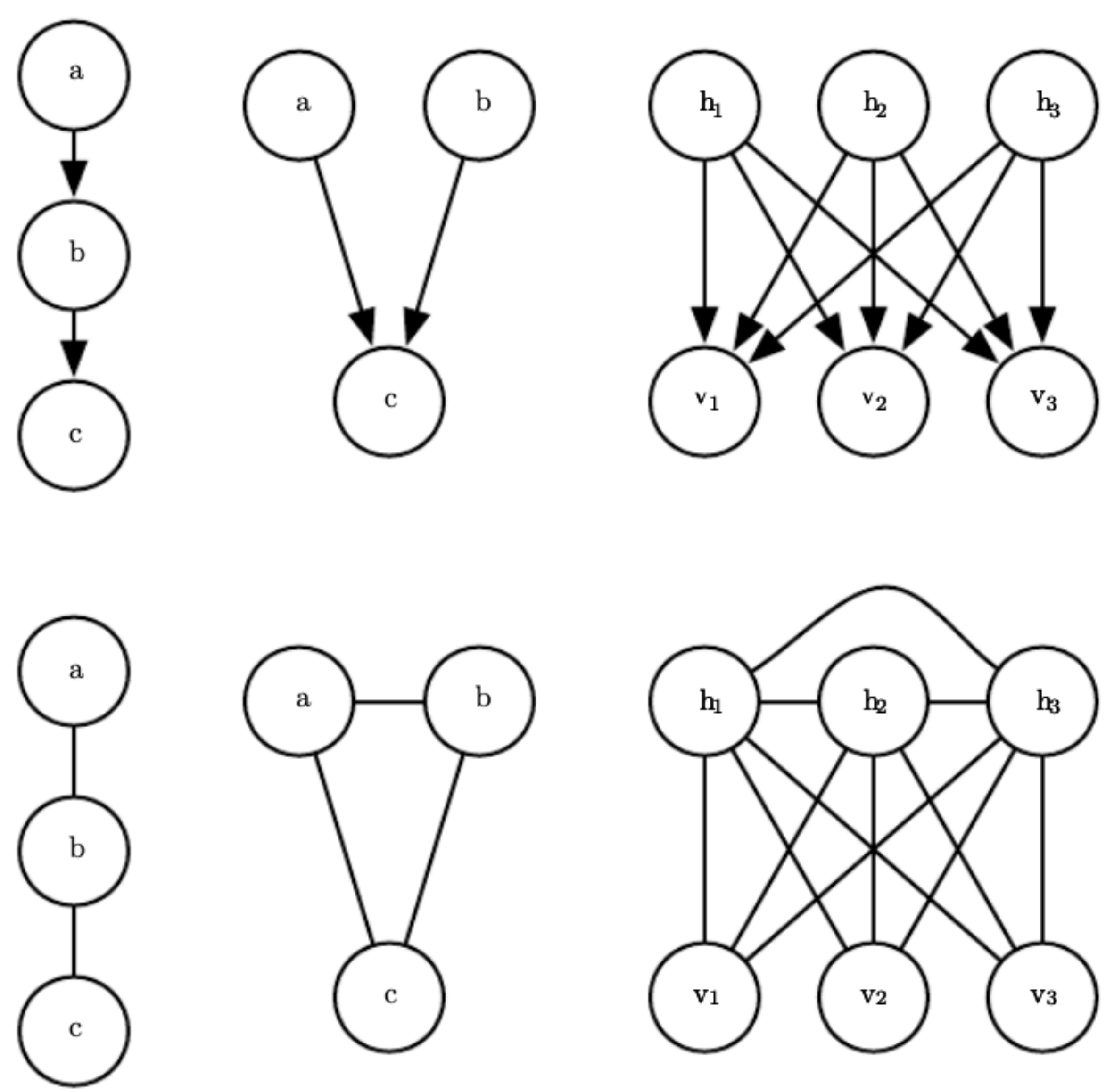
Undirected vs Directed Graphs

Converting directed models to undirected models (cannot be represented perfectly)

For every pair of variable x and y add an undirected edge (a moralised graph):

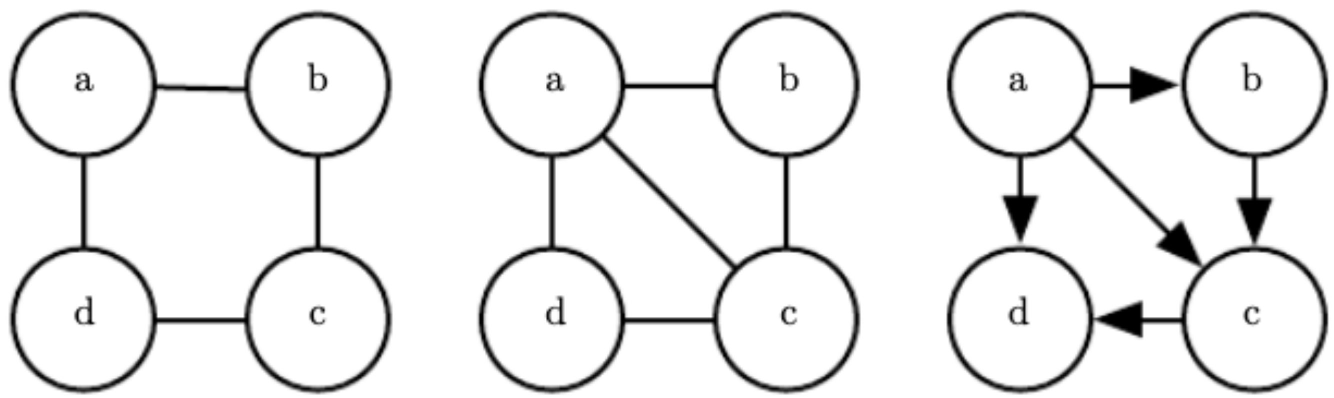
if there is a directed edge, or,

if x and y are parents of a node.



Undirected vs Directed Graphs

Similarly, undirected models can contain substructures that no directed model can represent (i.e., the latter cannot represent all conditional independencies in the former)



In conclusion: directed and undirected graphs encode strictly different (conditional) independence information

Deep Learning Book, Ian Goodfellow et al., page 568

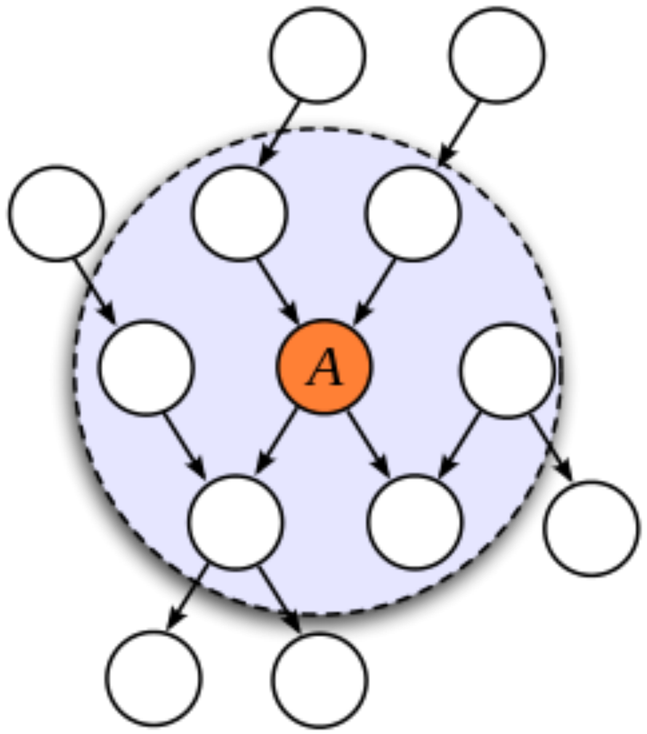
Markov blanket & boundary

A **Markov blanket** of a random variable Y in a random variable set $\mathcal{S} = \{X_1, \ldots, X_n\}$ is any subset $\mathcal{S}_1$ of $\mathcal{S}$, conditioned on which other variables are independent with Y:

$$Y \perp\!\!\!\perp \mathcal{S} \backslash \mathcal{S}_1 \mid \mathcal{S}_1 .$$

It means that $\mathcal{S}_1$ contains all the information one needs to infer Y, and the variables in $\mathcal{S} \backslash \mathcal{S}_1$ are redundant.

In general, Markov blanket is not unique. Any set in $\mathcal{S}$ that contains a Markov blanket is also a Markov blanket itself.



A **Markov boundary** is a Markov blanket none of whose subsets are Markov blankets themselves

parents of A, children of A,
parents of children of A

Wikipedia/Markov_Blanket

Causal Discovery

Learning causal relationships: Learn set of edges

- A causal structure **constrains** the possible types of probability distribution that can be generated from that structure.

Learning causal relationships: Learn set of edges

- A causal structure **constrains** the possible types of probability distribution that can be generated from that structure.
- Reverse: Obtain causal structures from probability distributions via causal inference

Learning causal relationships: Learn set of edges

- A causal structure **constrains** the possible types of probability distribution that can be generated from that structure.
- Reverse: Obtain causal structures from probability distributions via causal inference
- Types of constraints: **Conditional independencies** (all parametric distributions), Vanishing determinants of partial covariance matrices (linear Gaussian with unobserved confounders), **Unequal dependence on residuals** (Non-linear additive noise, or linear non-Gaussian), **interventions/ perturbations**, time-series …

Causal Discovery Methods (Based on Graphical Models)

Class of Algorithm	Name	Assumptions	Short comings	Input
Constraint-based	PC (oldest)	Any distribution, No unobsv. confounders, Markov cond, faithfulness	Causal info only up to equivalence classes, Non bivariate	Complete undirected graph
	FCI	Any distribution, Asymptotically correct with confounders, Markov cond, faithfulness		
Score-based	GES	No unobsv. confounders	Non-bivariate	Empty graph, adds edges, removes some
Functional Causal Models (FCMs)	LinGAM/ ANM	Asymmetry in data	Requires additional assumptions (not general), harder for discrete data	Structural Equation Model

Glymour et al. (2019)

Assumptions 1: The Markov Condition

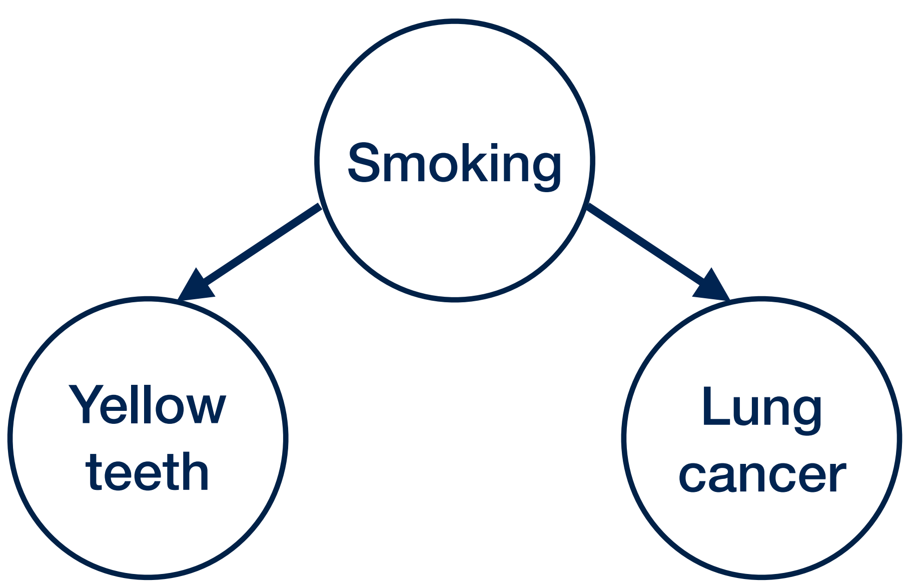
Any variable X is independent of all other variables, conditional on its parents (PA) and unobserved variables (noise):

$$P(x_1, \cdots, x_n) = \prod_{j=1}^{J} P(x_j | PA_j, \epsilon_j)$$

if conditional dependence, there is an edge

- Absent edge implies conditional independence (**CI**)
- Observing conditional dependence implies an edge

For example: Yellow teeth, lung cancer, smoking



An edge is wrongly inferred, when parent is omitted

Assumptions 2 & 3: Causal sufficiency & Faithfulness

- **Causal sufficiency:** For any pair of variables X, Y, if there exists a variable Z which is a direct of cause of both X and Y, then Z is included in the causal graph (Z may be unobserved)

- A probability distribution P is **faithful** to a DAG G if no CI relations other than the ones entailed by the Markov property are present.
  - Conjugate to the Markov condition
  - Edge implies conditional dependence
  - Observing CI implies absence of an edge

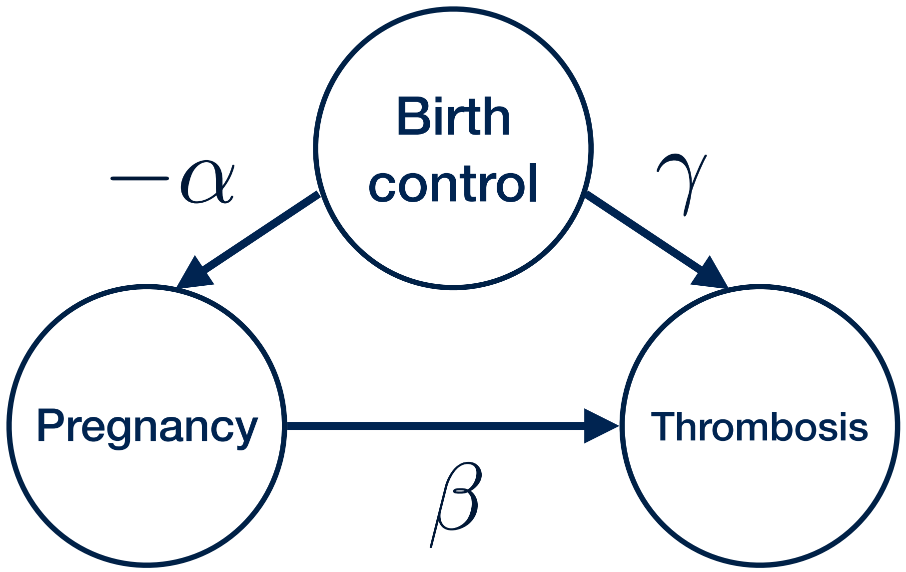
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



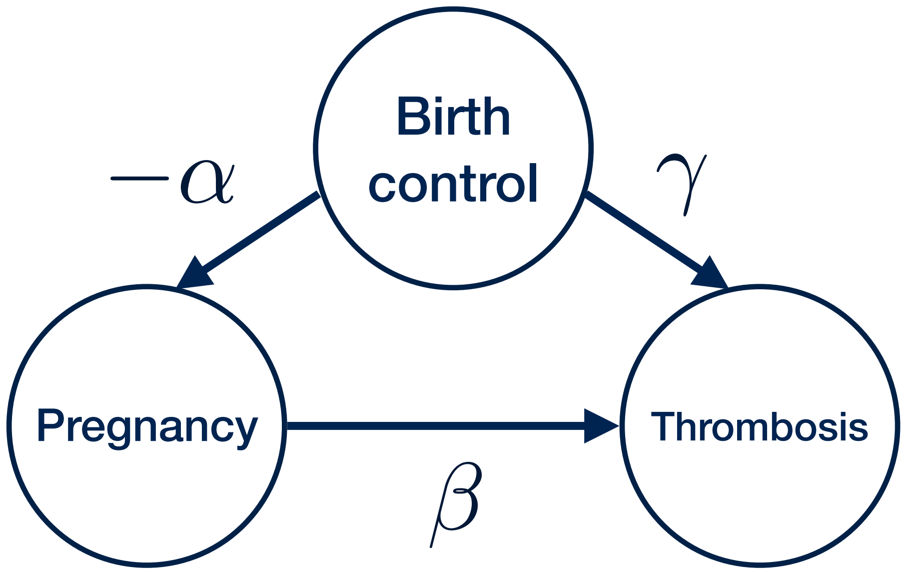
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



So if $\gamma = \alpha\beta$, no dependency between T and B will be observed!

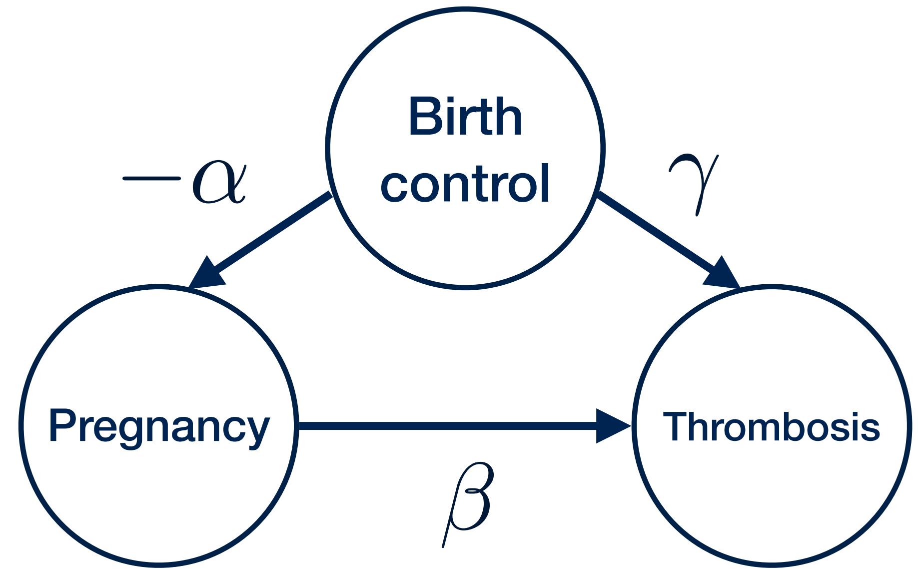
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



So if $\gamma = \alpha\beta$, no dependency between T and B will be observed!

- Fails in **regulatory systems**, e.g. home temperature, outside temp, thermostat: By design, thermostat keeps the inside temp independent of outside, always fixed at T*

in real world, conditional independency doesn't mean delete a edge

Silver 2018

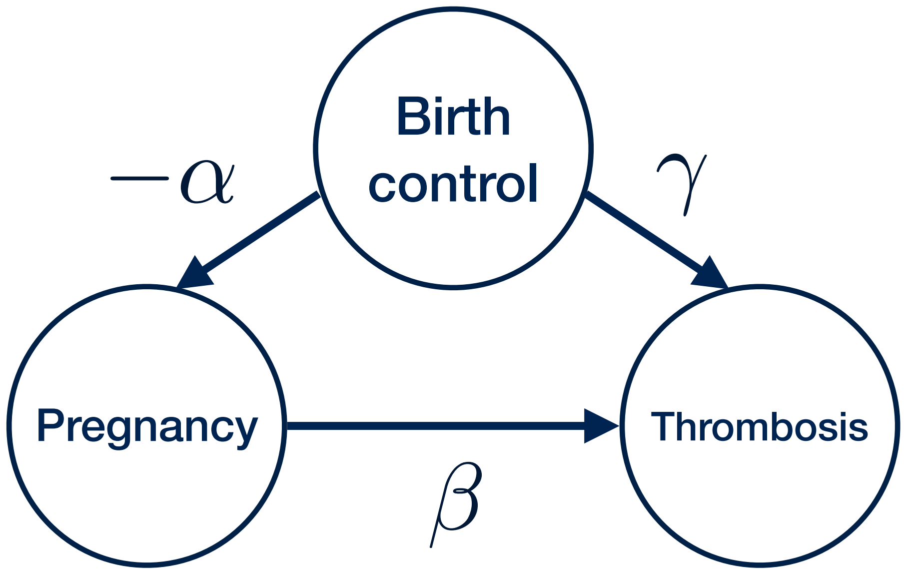
Assumptions 3: Faithfulness

It **fails** when distributions are set up in such a way that paths exactly cancel:

$$P = -\alpha B + U_P$$

$$T = \beta P + \gamma B + U_T$$

$$\Rightarrow T = (-\alpha\beta + \gamma)B + U$$



So if $\gamma = \alpha\beta$, no dependency between T and B will be observed!

- Fails in **regulatory systems**, e.g. home temperature, outside temp, thermostat: By design, thermostat keeps the inside temp independent of outside, always fixed at T*
- **Biology and homeostasis!**

Often keep the assumption and argue that most distributions are multimodal and will not cancel each other exactly …

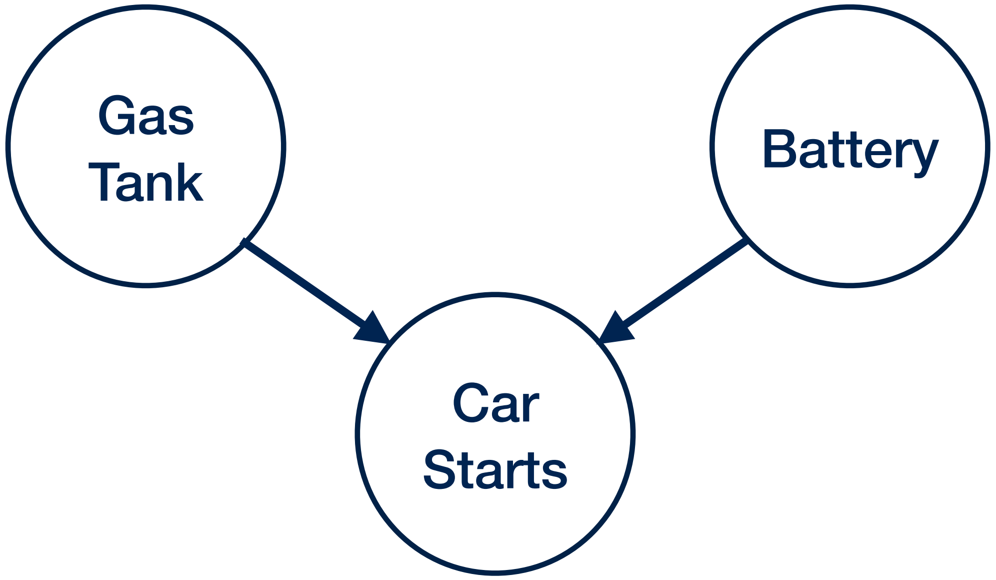
Silver 2018

Distinguishing causal structures: V-structures

- Recall collider example:

Gas tank $\perp\!\!\!\perp$ Battery

Gas tank $\not\perp\!\!\!\perp$ Battery | Car starts = 0

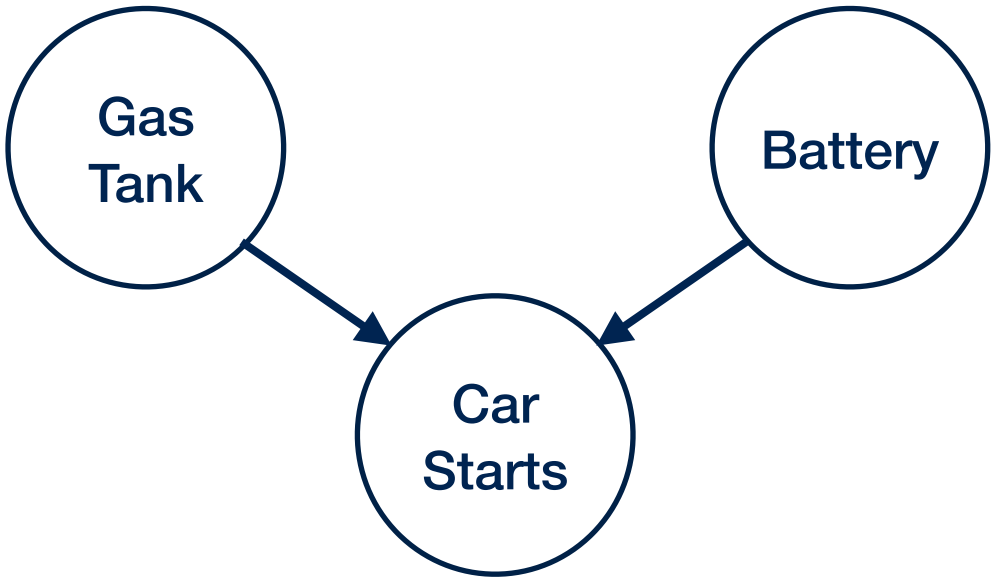


Distinguishing causal structures: V-structures

- Recall collider example:

Gas tank $\perp\!\!\!\perp$ Battery

Gas tank $\not\perp\!\!\!\perp$ Battery | Car starts = 0



- **Markov Equivalence Class (MEC)**: Two graphs G and G' belong to the same equivalence class iff each conditional independence implied by G is also implied by G' and vice versa.

- We can learn edges/directions using MEC and d-separation.
- D-separations gives all CI implied by graph

Markov Equivalence Class (MEC)

True DAG	$A \rightarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$
All Observed CIs	$A \perp\!\!\!\perp C \mid B$	$A \perp\!\!\!\perp C \mid \emptyset$
Set of DAGs in MEC	$A \rightarrow B \rightarrow C$ $A \leftarrow B \leftarrow C$ $A \leftarrow B \rightarrow C$	$A \rightarrow B \leftarrow C$
CPDAG (complete partially DAG)	$A - B - C$	$A \rightarrow B \leftarrow C$

Glymour et al. (2019)

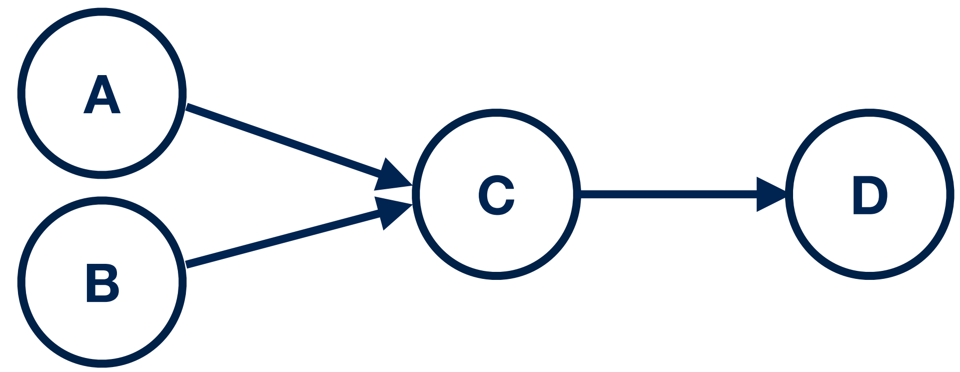
The Search Space of Causal Graphs

- For |V|=n nodes there are $\binom{n}{2} = \frac{1}{2}(n-1)n$ distinct pairs of variables
- There are at least $2^{\frac{1}{2}(n-1)n}$ possible graphs where between any two pairs there is either an edge or no edge.
- There are at most $3^{\frac{1}{2}(n-1)n}$ possible graphs since we may have either of: $A \rightarrow B, \; A \leftarrow B, \; A \qquad B$
- Grows super exponentially in the number of nodes
- Requires efficient causal discovery algorithms: PC algorithm

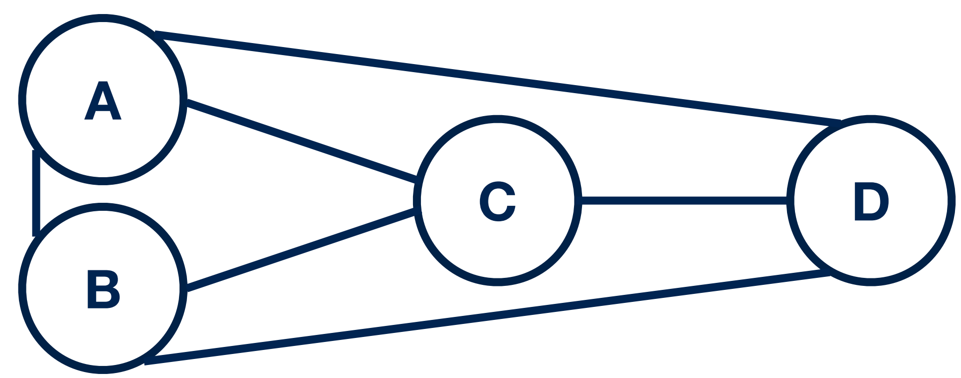
Silver 2018

Peter-Clark (PC) Algorithm

True causal graph:

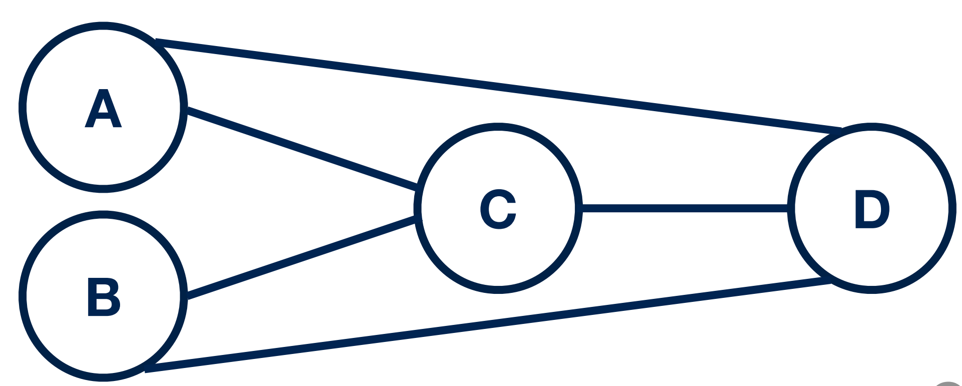


1. Start with the complete graph



2. Zeroth order CI, $A \perp\!\!\!\perp B$, by faithfulness:

Need statistical independence testing.

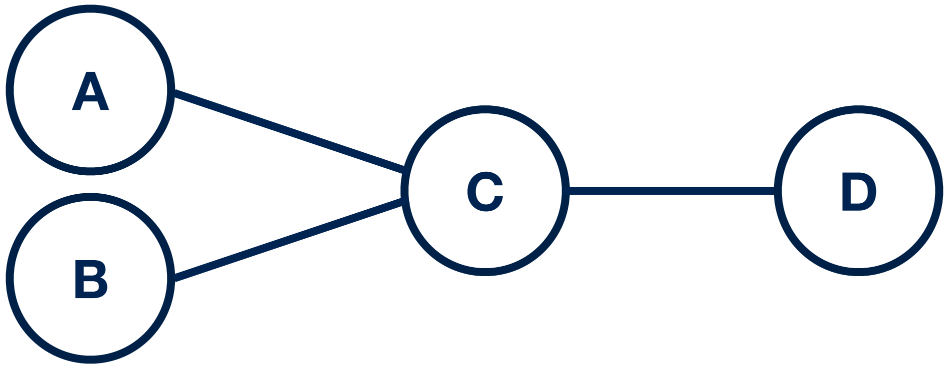


Glymour et al. (2019)

Peter-Clark (PC) Algorithm

3. 1st order CI, $A \perp\!\!\!\perp D | C$, by faithfulness:

$B \perp\!\!\!\perp D | C$

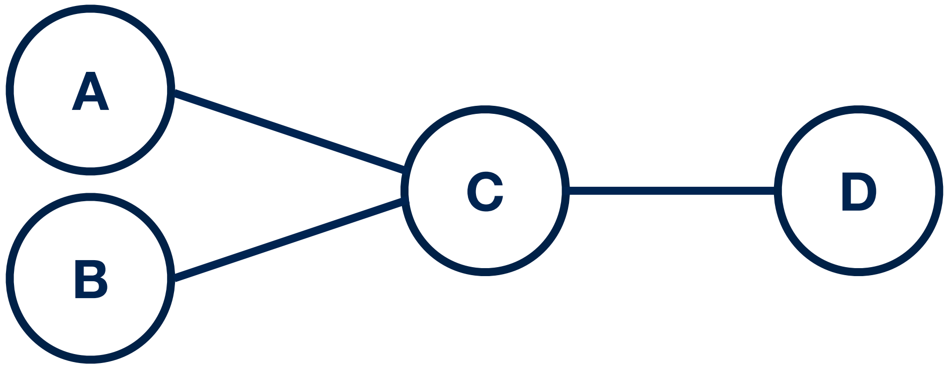


Glymour et al. (2019)

Peter-Clark (PC) Algorithm

3. 1st order CI, $A \perp\!\!\!\perp D|C$, by faithfulness:

$$B \perp\!\!\!\perp D|C$$



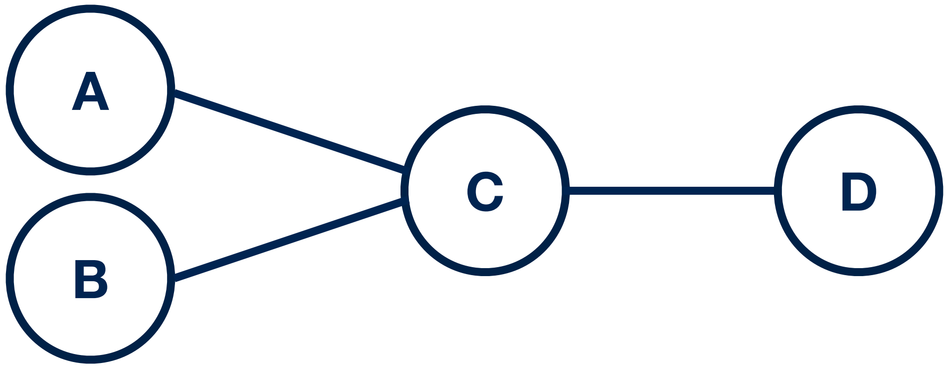
4. No higher order CI observed. Notice that conditioning sets only need to contain **neighbours** for the two nodes due to the Markov condition. We do not know the parents but parents are a subsets of neighbours. As the graph becomes sparser, the number of tests to be performed decreases. This makes PC very efficient.

Glymour et al. (2019)

Peter-Clark (PC) Algorithm

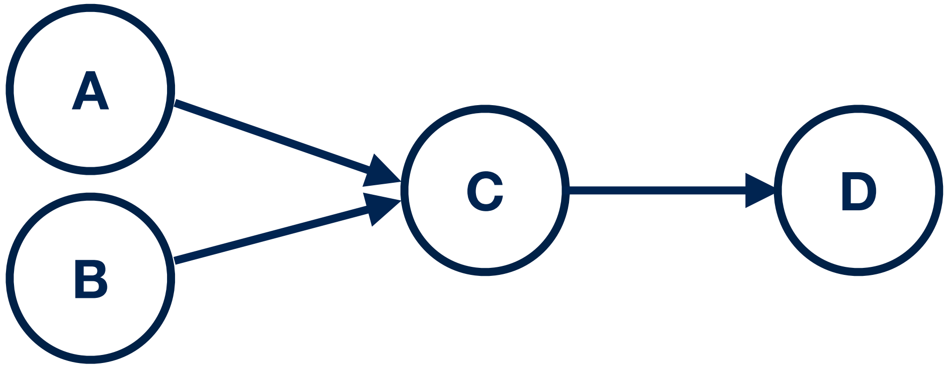
3. 1st order CI, $A \perp\!\!\!\perp D | C$, by faithfulness:

$B \perp\!\!\!\perp D | C$



4. No higher order CI observed. Notice that conditioning sets only need to contain **neighbours** for the two nodes due to the Markov condition. We do not know the parents but parents are a subsets of neighbours. As the graph becomes sparser, the number of tests to be performed decreases. This makes PC very efficient.

5. Orient V-structures (colliders): take triplets where 2 nodes are connected to the 3rd: $A \not\!\perp\!\!\!\perp B | C$ only.



Note $C \leftarrow D$ cannot be as it would have been a collider (not detected in 5)

Glymour et al. (2019)

Remarks

- Missing/unobserved variables could lead to wrong/biased graphs
- Conditional independence tests are subject of active research
- Parallelised PC
- PC for heterogeneous data etc.
- PC + score-based

Methods for Causal Inference
Lecture 12

Ava Khamseh
School of Informatics

2021-2022