SEM 2: Structural Equation Modeling

Week 3 - Causality & DAGs

Sacha Epskamp

Given the following causal statement:

Exercising leads to a higher heart rate

Given the following causal statement:

Exercising leads to a higher heart rate

Which of the below statements are plausible?

- If you are not exercising, your heart rate must be low

Given the following causal statement:

Exercising leads to a higher heart rate

Which of the below statements are plausible?

- If you are not exercising, your heart rate must be low
- If you are exercising, your heart rate must be high

Given the following causal statement:

Exercising leads to a higher heart rate

Which of the below statements are plausible?

- If you are not exercising, your heart rate must be low
- If you are exercising, your heart rate must be high
- If you are exercising, your heart rate is likely to be higher than if you are not exercising

Given the following causal statement:

Exercising leads to a higher heart rate

Which of the below statements are plausible?

- If you are not exercising, your heart rate must be low
- If you are exercising, your heart rate must be high
- If you are exercising, your heart rate is likely to be higher than if you are not exercising
- If your heart rate is high, you must be exercising

Given the following causal statement:

Exercising leads to a higher heart rate

Which of the below statements are plausible?

- If you are not exercising, your heart rate must be low
- If you are exercising, your heart rate must be high
- If you are exercising, your heart rate is likely to be higher than if you are not exercising
- If your heart rate is high, you must be exercising
- If we see that you are exercising, we can predict your heart rate to be higher than average

Given the following causal statement:

Exercising leads to a higher heart rate

Which of the below statements are plausible?

- If you are not exercising, your heart rate must be low
- If you are exercising, your heart rate must be high
- If you are exercising, your heart rate is likely to be higher than if you are not exercising
- If your heart rate is high, you must be exercising
- If we see that you are exercising, we can predict your heart rate to be higher than average
- If we see that you have a high heart rate, we can predict that you are exercising

$\mathrm{Exercising} \rightarrow$ Higher heart rate

Implies:

- Observing that you are exercising makes it more likely that you have a higher heart rate
  - $\mathcal{E}(\text{heart rate} \mid \mathrm{See}(\text{exercising})) > \mathcal{E}(\text{heart rate})$

$\mathrm{Exercising} \rightarrow$ Higher heart rate

Implies:

- Observing that you are exercising makes it more likely that you have a higher heart rate
  - $\mathcal{E}(\text{heart rate} \mid \mathrm{See}(\text{exercising})) > \mathcal{E}(\text{heart rate})$
- Observing that you have a high heartrate makes it more likely you are exercising a lot
  - $\mathcal{E}(\text{exercising} \mid \mathrm{See}(\text{high heart rate})) > \mathcal{E}(\text{exercising})$

$\mathrm{Exercising} \rightarrow$ Higher heart rate

Implies:

- Observing that you are exercising makes it more likely that you have a higher heart rate
  - $\mathcal{E}(\text{heart rate} \mid \mathrm{See}(\text{exercising})) > \mathcal{E}(\text{heart rate})$
- Observing that you have a high heartrate makes it more likely you are exercising a lot
  - $\mathcal{E}(\text{exercising} \mid \mathrm{See}(\text{high heart rate})) > \mathcal{E}(\text{exercising})$
- But *making* your heart rate high does not make you exercise!
  - $\mathcal{E}(\text{exercising} \mid \mathrm{Do}(\text{high heart rate})) = \mathcal{E}(\text{exercising})$

$\mathrm{Exercising} \rightarrow$ Higher heart rate

Implies:

- Observing that you are exercising makes it more likely that you have a higher heart rate
  - $\mathcal{E}(\text{heart rate} \mid \mathrm{See}(\text{exercising})) > \mathcal{E}(\text{heart rate})$
- Observing that you have a high heartrate makes it more likely you are exercising a lot
  - $\mathcal{E}(\text{exercising} \mid \mathrm{See}(\text{high heart rate})) > \mathcal{E}(\text{exercising})$
- But *making* your heart rate high does not make you exercise!
  - $\mathcal{E}(\text{exercising} \mid \mathrm{Do}(\text{high heart rate})) = \mathcal{E}(\text{exercising})$

Unfortunately, in observational data (especially without temporal ordering), we can only investigate what happens if we see one variable (conditioning)...

$\mathrm{Exercising} \rightarrow$ Higher heart rate

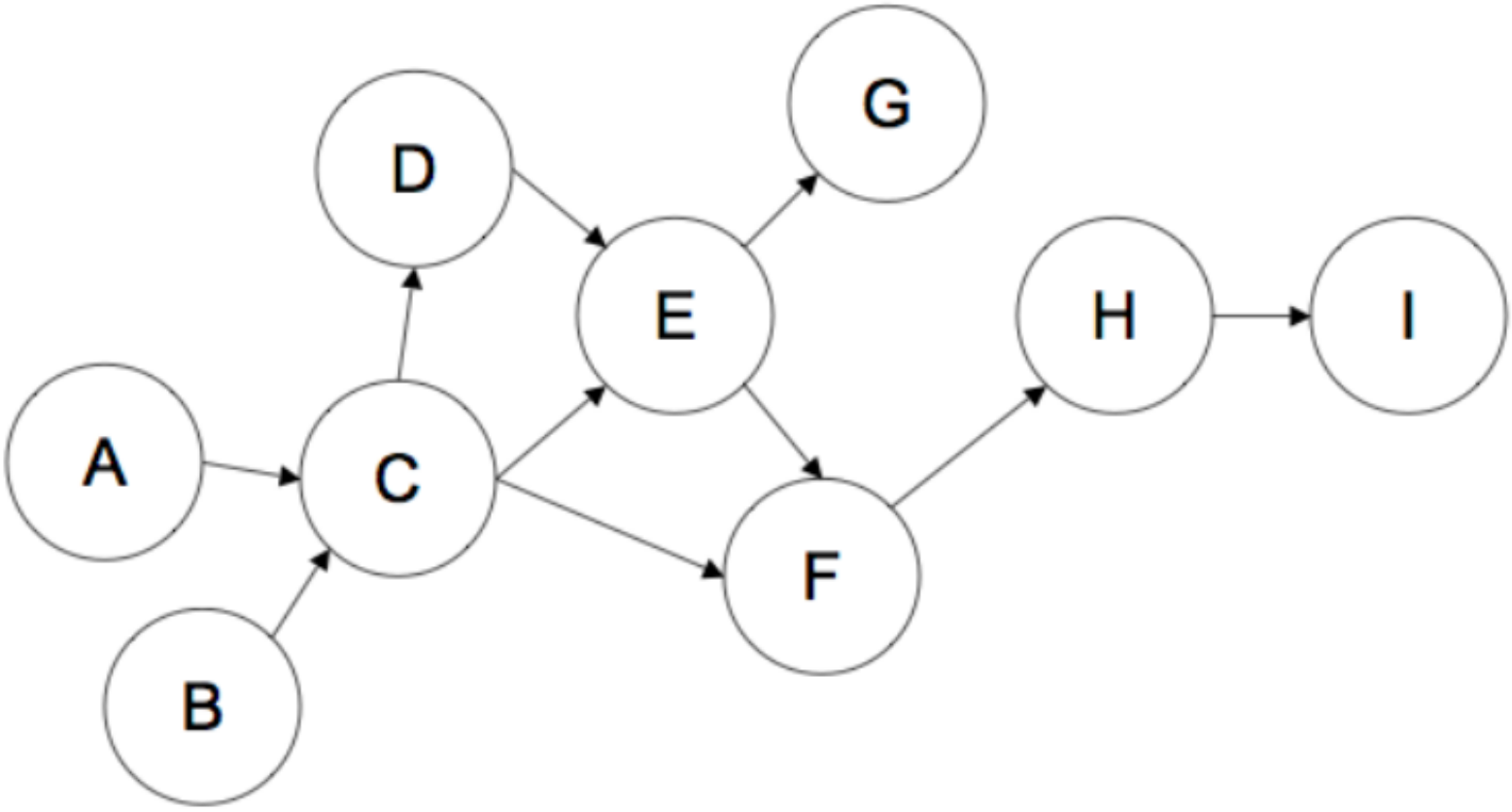
Implies:

- Observing that you are exercising makes it more likely that you have a higher heart rate
  - $\mathcal{E}(\text{heart rate} \mid \mathrm{See}(\text{exercising})) > \mathcal{E}(\text{heart rate})$
- Observing that you have a high heartrate makes it more likely you are exercising a lot
  - $\mathcal{E}(\text{exercising} \mid \mathrm{See}(\text{high heart rate})) > \mathcal{E}(\text{exercising})$
- But *making* your heart rate high does not make you exercise!
  - $\mathcal{E}(\text{exercising} \mid \mathrm{Do}(\text{high heart rate})) = \mathcal{E}(\text{exercising})$

Unfortunately, in observational data (especially without temporal ordering), we can only investigate what happens if we see one variable (conditioning)...

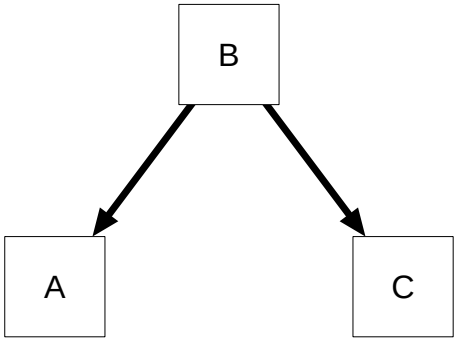
Solution: More variables and more advanced causal models imply more testable hypotheses (conditional independence relations)!

Directed Acyclic Graphs



Building blocks of a DAG

Common Cause

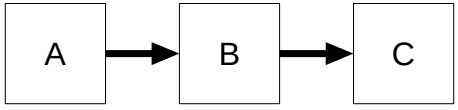


Example: Disease (B) causes two symptoms (A and C).

$A \not\perp\!\!\!\perp C$

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

Chain

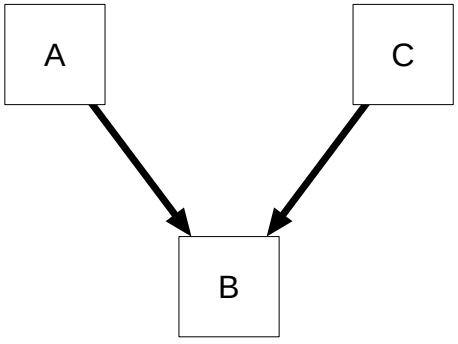


Example: Insomnia (A) causes fatigue (B), which in turn causes concentration problems (C)

$A \not\perp\!\!\!\perp C$

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

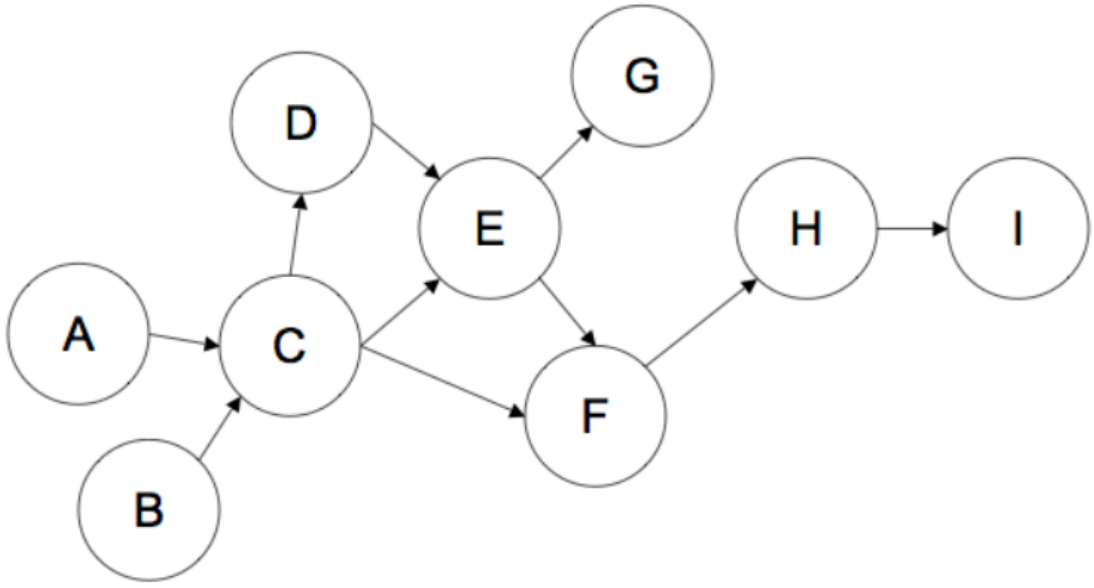
Collider



Example: Difficulty of class (A) and motivation of student (C) cause grade on a test (B)

$A \perp\!\!\!\perp C$

$A \not\perp\!\!\!\perp C \mid B$



- $A \perp\!\!\!\perp B$
- $A \perp\!\!\!\perp D \mid C$
- $B \perp\!\!\!\perp G \mid C, E$
- ...

Testing this causal model involves testing if all these conditional independence relations hold