A Brief Introduction to Causal Inference and Causal Diagrams

Alireza Akhondi-Asl
MSICU Center For Outcomes
Department of Anesthesiology, Critical Care and Pain Medicine

Learning Objectives

Causal Diagrams / Directed Acyclic Graphs (DAGs)	Our World Model
Conditional Dependency/Independency in Causal Graphs	Statistical implications of the model
Identification of Causal Effects from DAGs	Using Observational data for causal inference

Causal Inference

Reasoning about the causal effect of a treatment

Potential Outcome

Outcome under a potential treatment.

What might have occurred under different treatments.

Causal Effect

Difference between the potential outcome when the treatment is received and potential outcome when the treatment is not received.

Fundamental limitation of Causal Inference

We observe only a potential outcome.

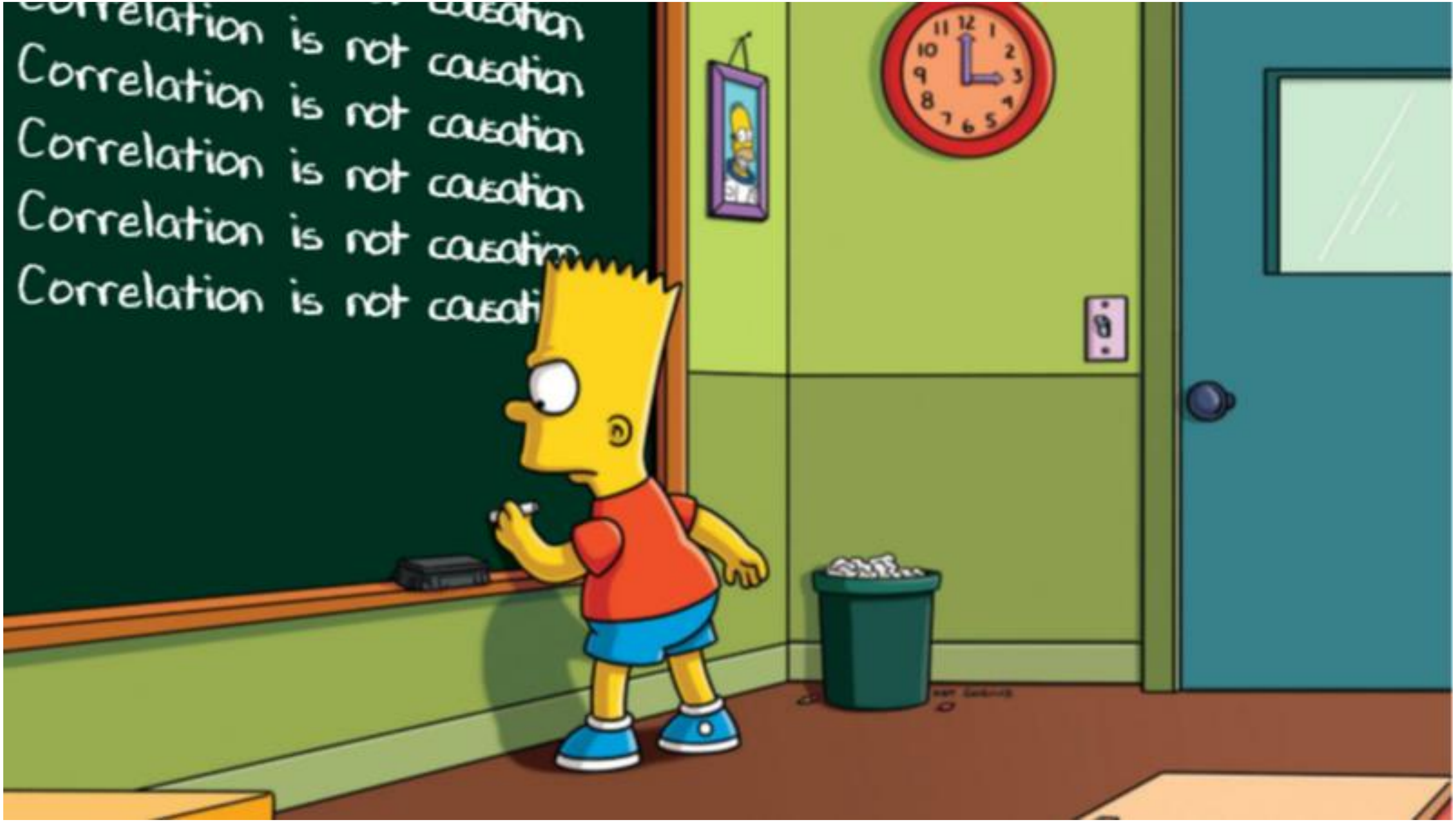
RCT vs Observational Study

Randomized Control Trial (RCT)

- All factors are random except the treatment
- Any change in the outcome is due to treatment (Causal Effect)

Why Observational Studies?

- Unethical
- Impractical
- Impossible
- Data is available


Correlation is not causation
Correlation is not causation
Correlation is not causation
Correlation is not causation
Correlation is not causation

Observational Studies

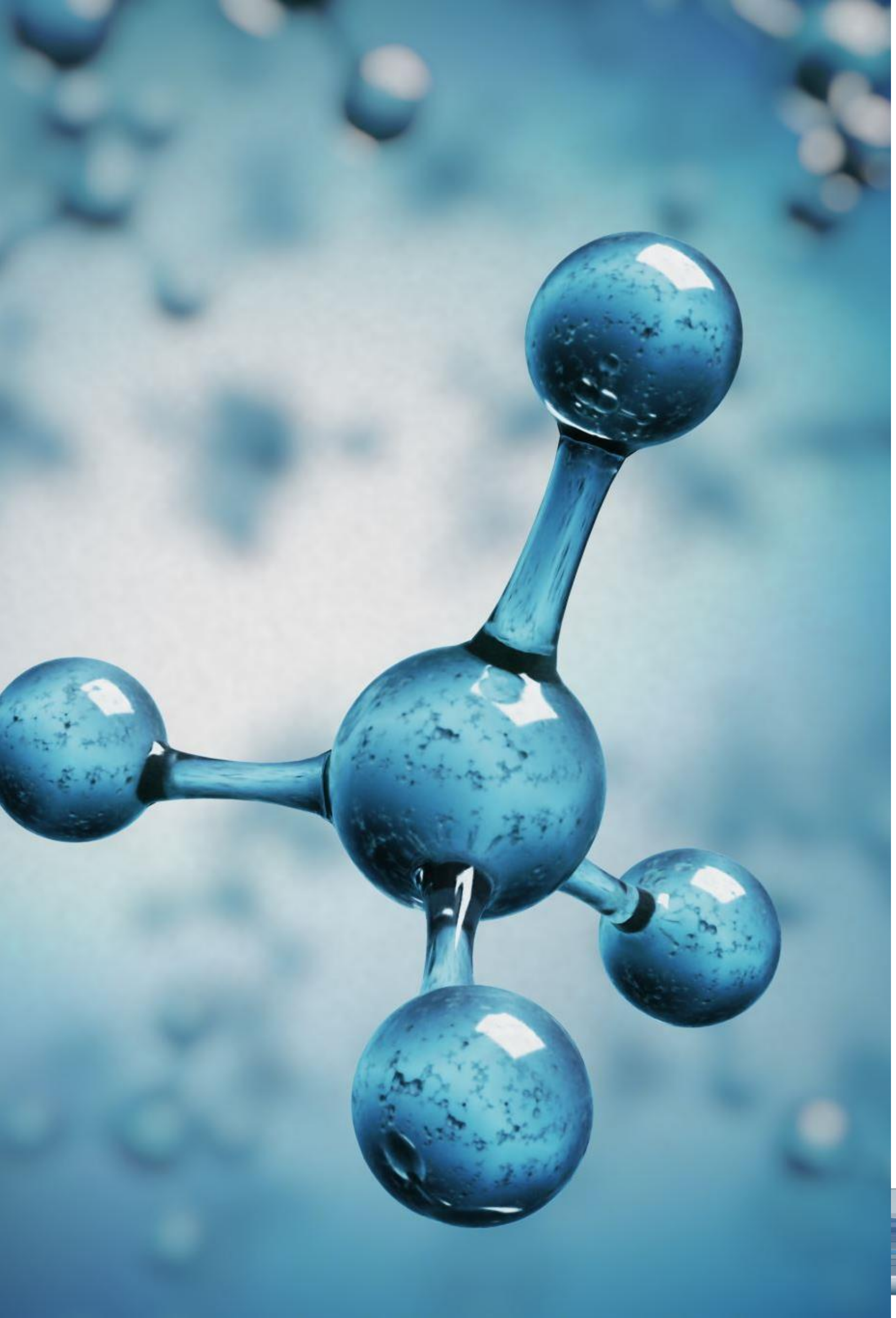
Treatment selection is influenced by subject characteristics.

Baseline characteristics are systematically different.

We should account for it when we are estimating the treatment effect.

If we know the data generation model,

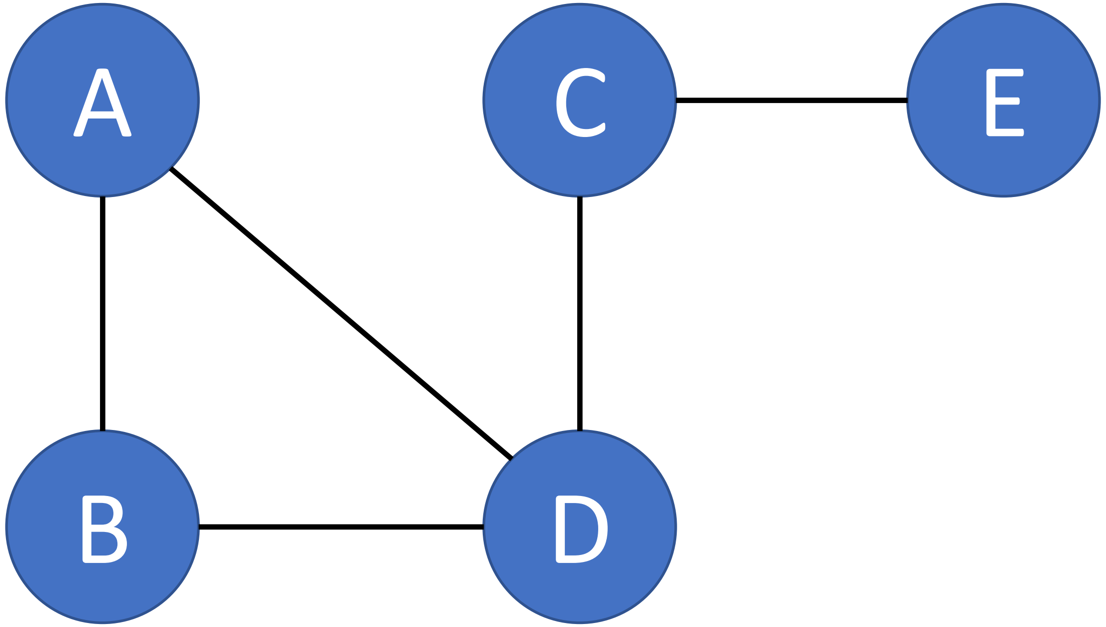
we might be able to identify causal effect from observational data!



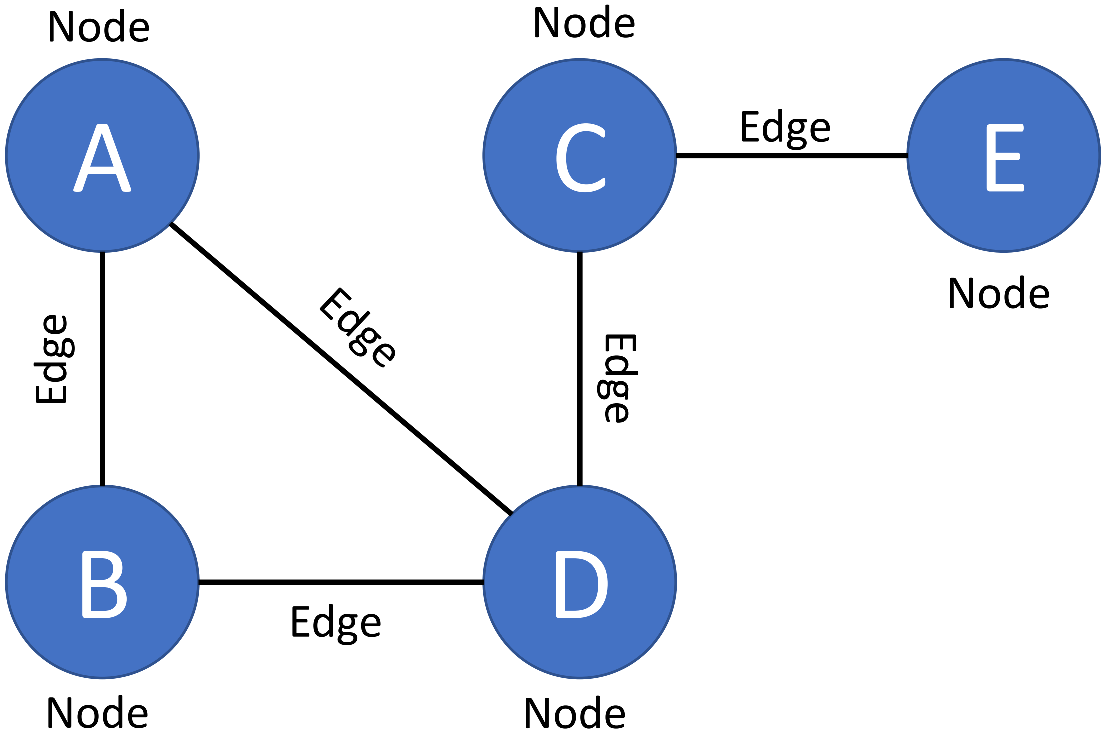
Structural Causal Model (SCM)

- Describes our **<u>assumptions</u>** about the relevant features of the world and the interaction of these features.
  - How variables are assigned
  - ***<u>If our assumptions are wrong, the model will be wrong</u>***
- Causal effect from observational data
- Every SCM is associated with a DAG

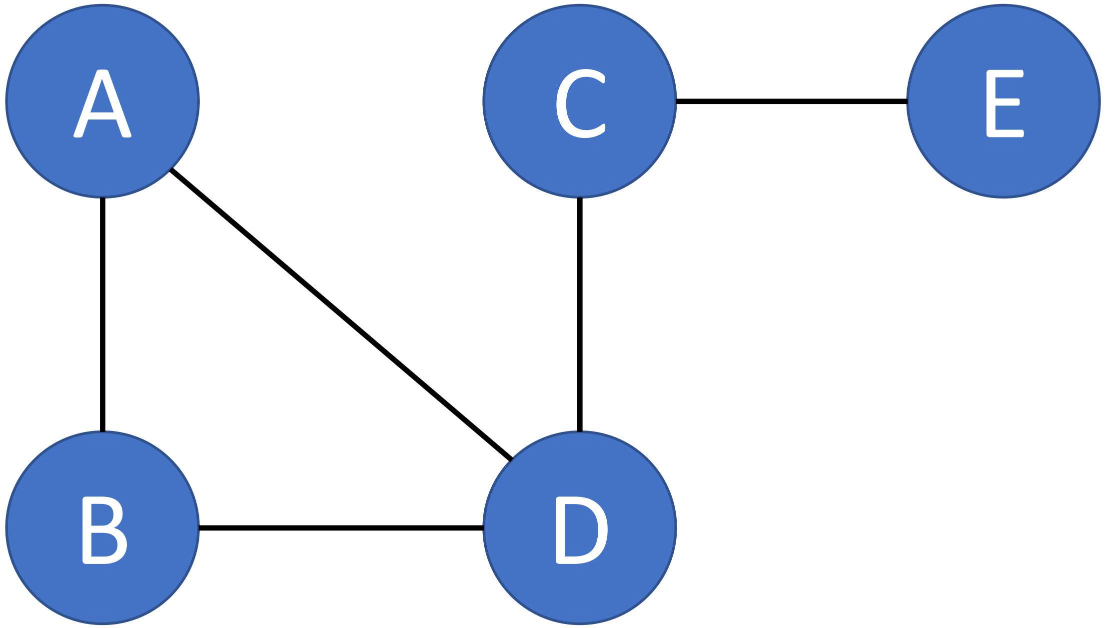
Graphs



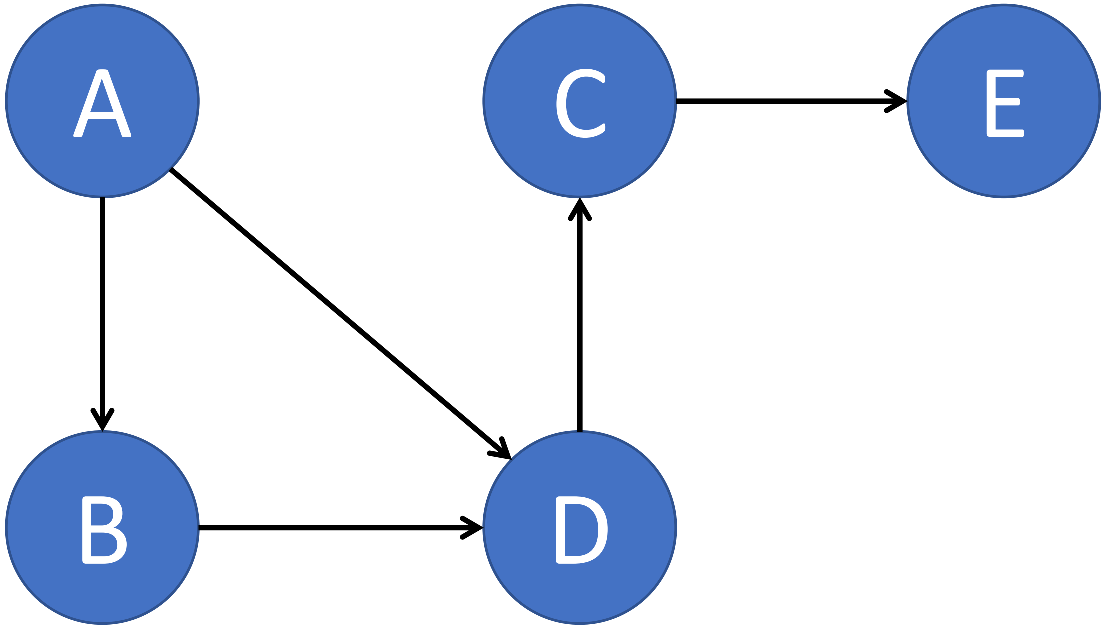
Nodes and Edges



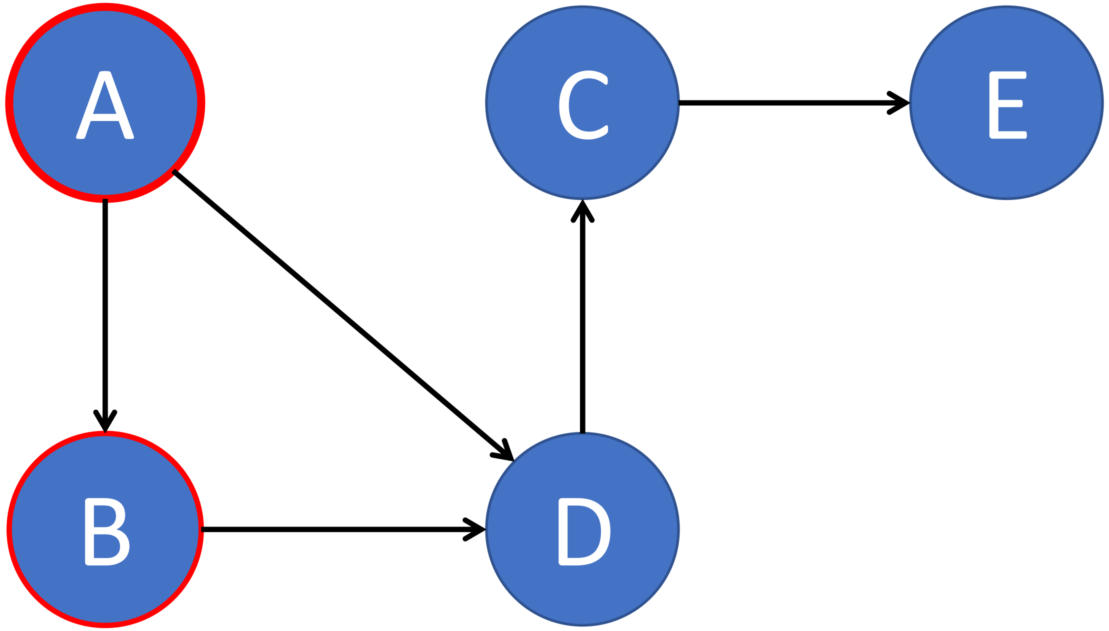
Undirected Graph



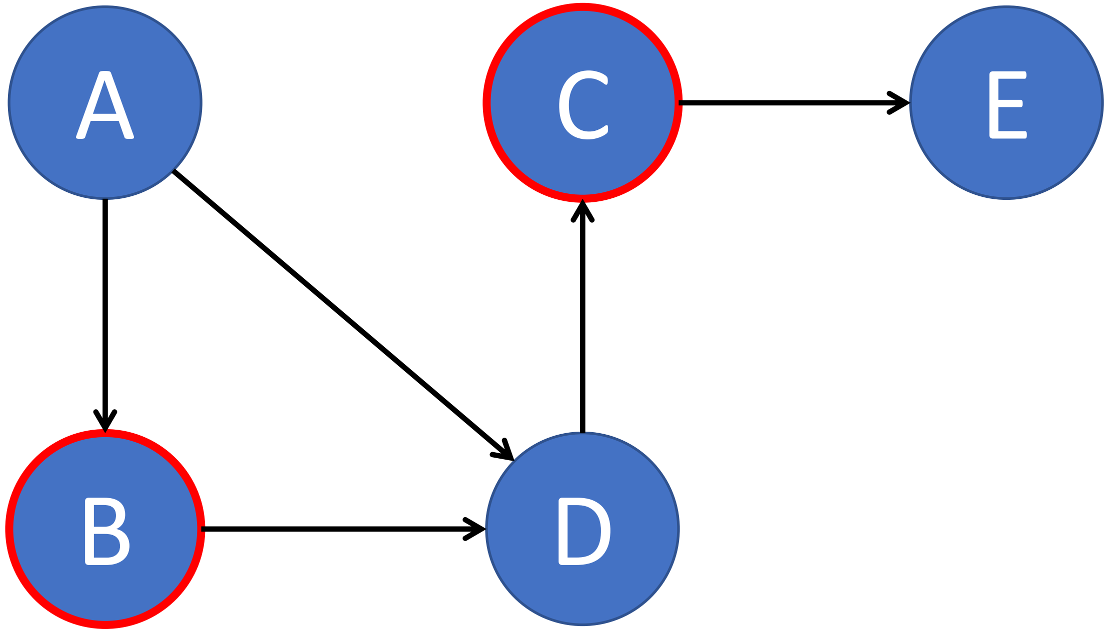
Directed Graph



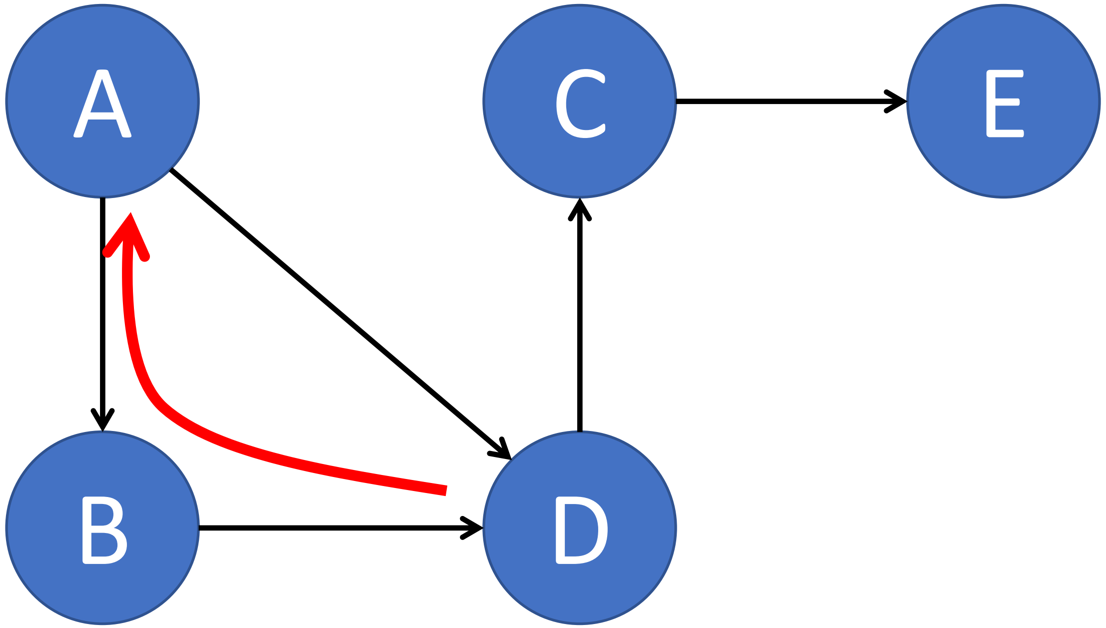
Adjacent Nodes



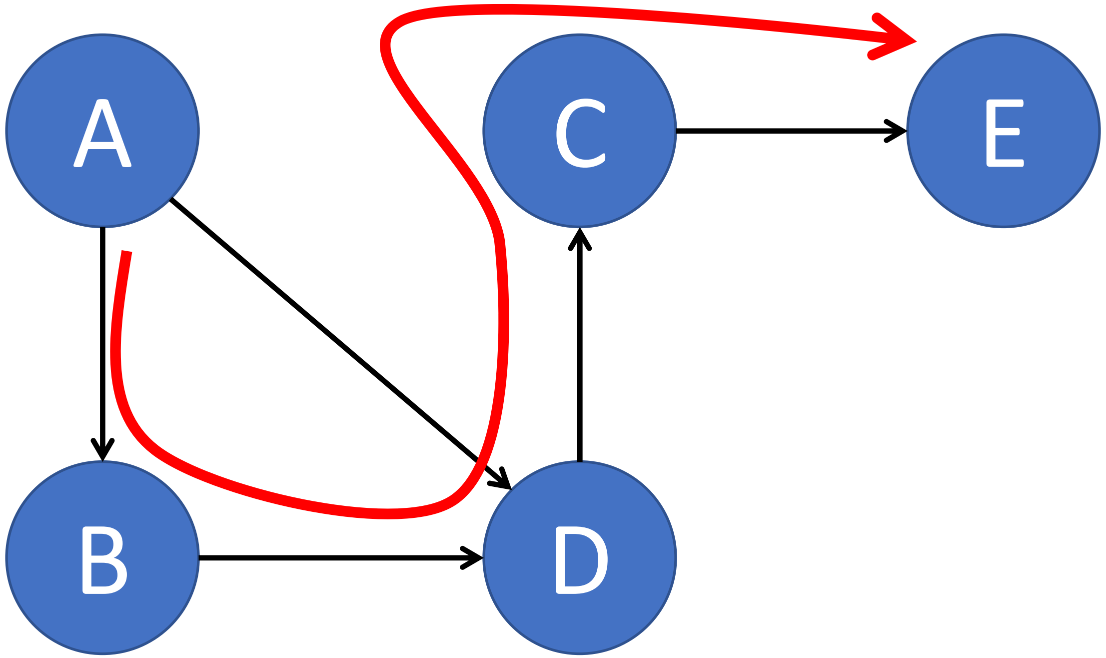
Not Adjacent Nodes



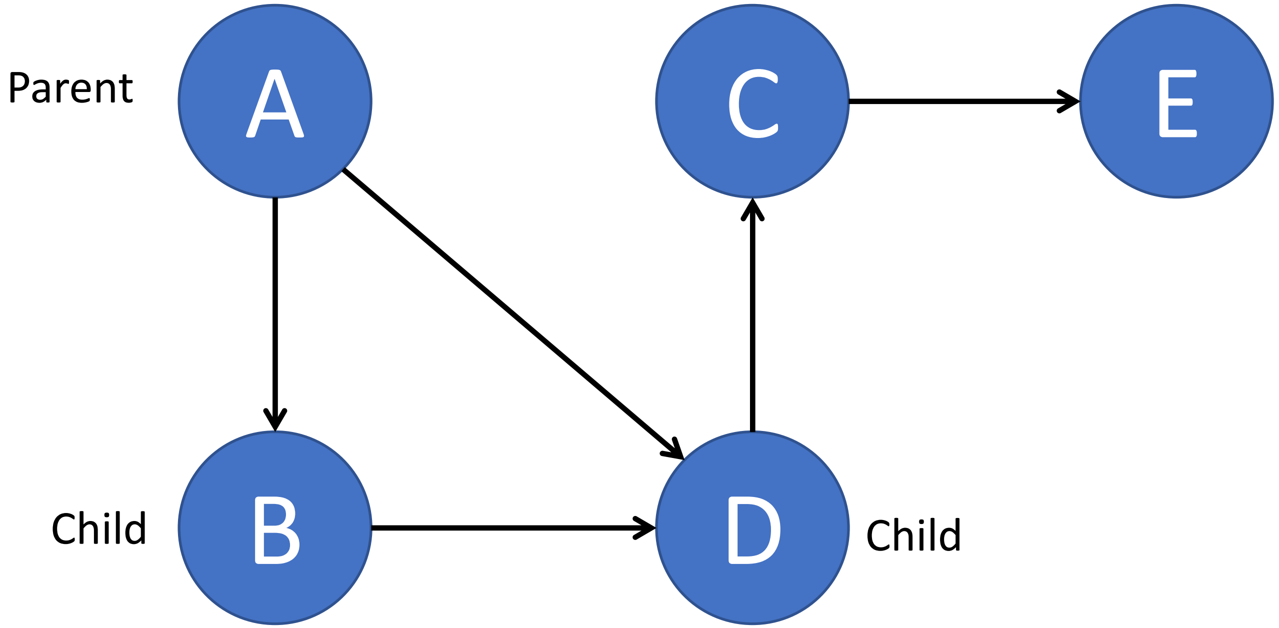
Path



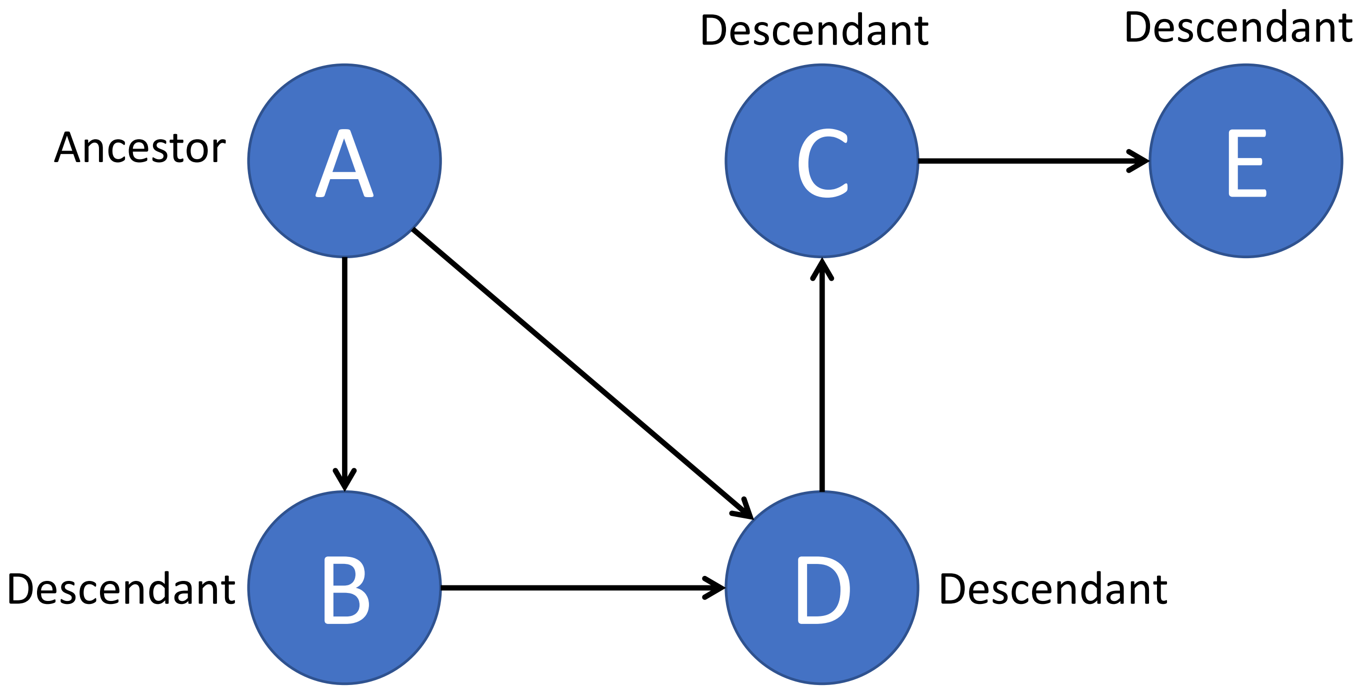
Directed Path



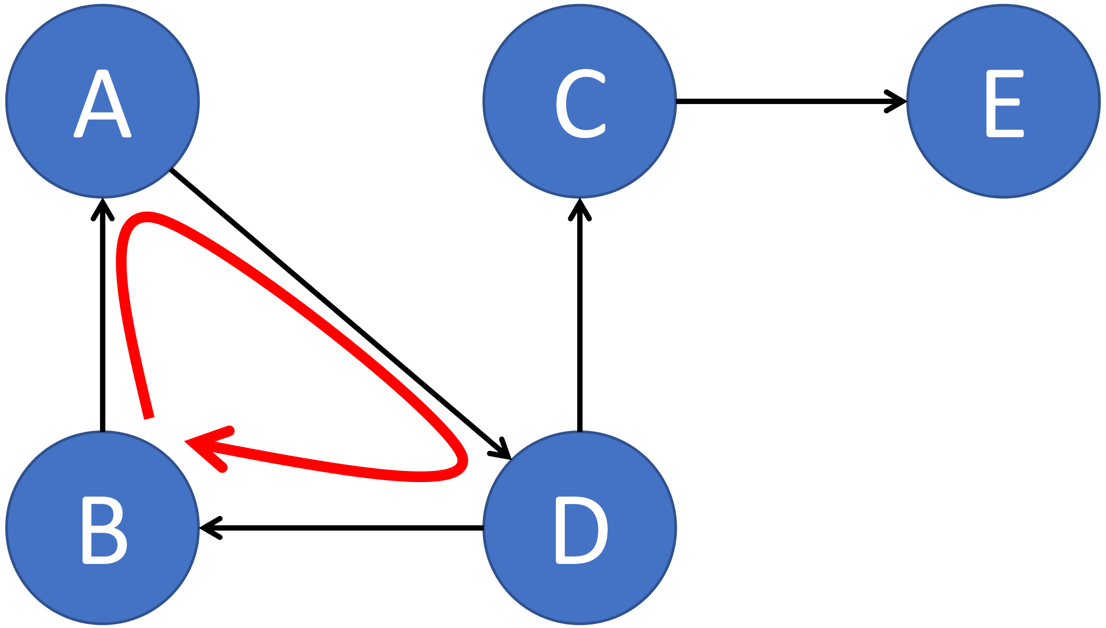
Parent - Child



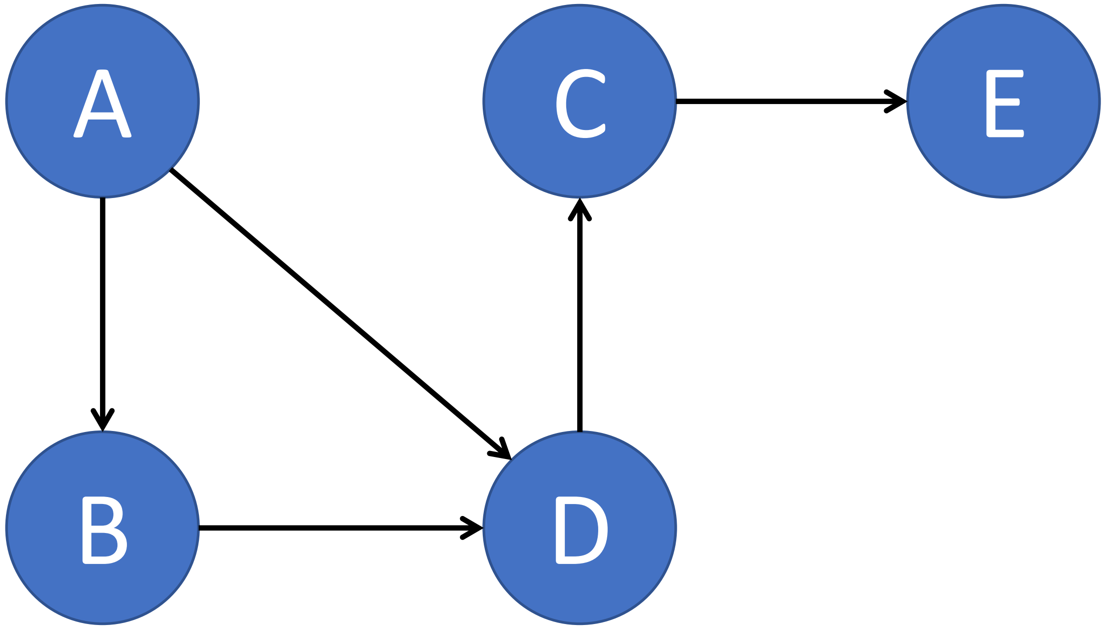
Ancestor - Descendant



Cycle



Directed Acyclic Graph (DAG)



DAGs

Graphically show the assumed data generation process

A blend of Structural equation modeling and Bayesian Networks

Well-matched with potential outcomes framework of causality

One of the main framework of causal inference

DAGs

Nonparametric

- No assumption about the form of the function and distribution

Intuitive

Strong Mathematical Support

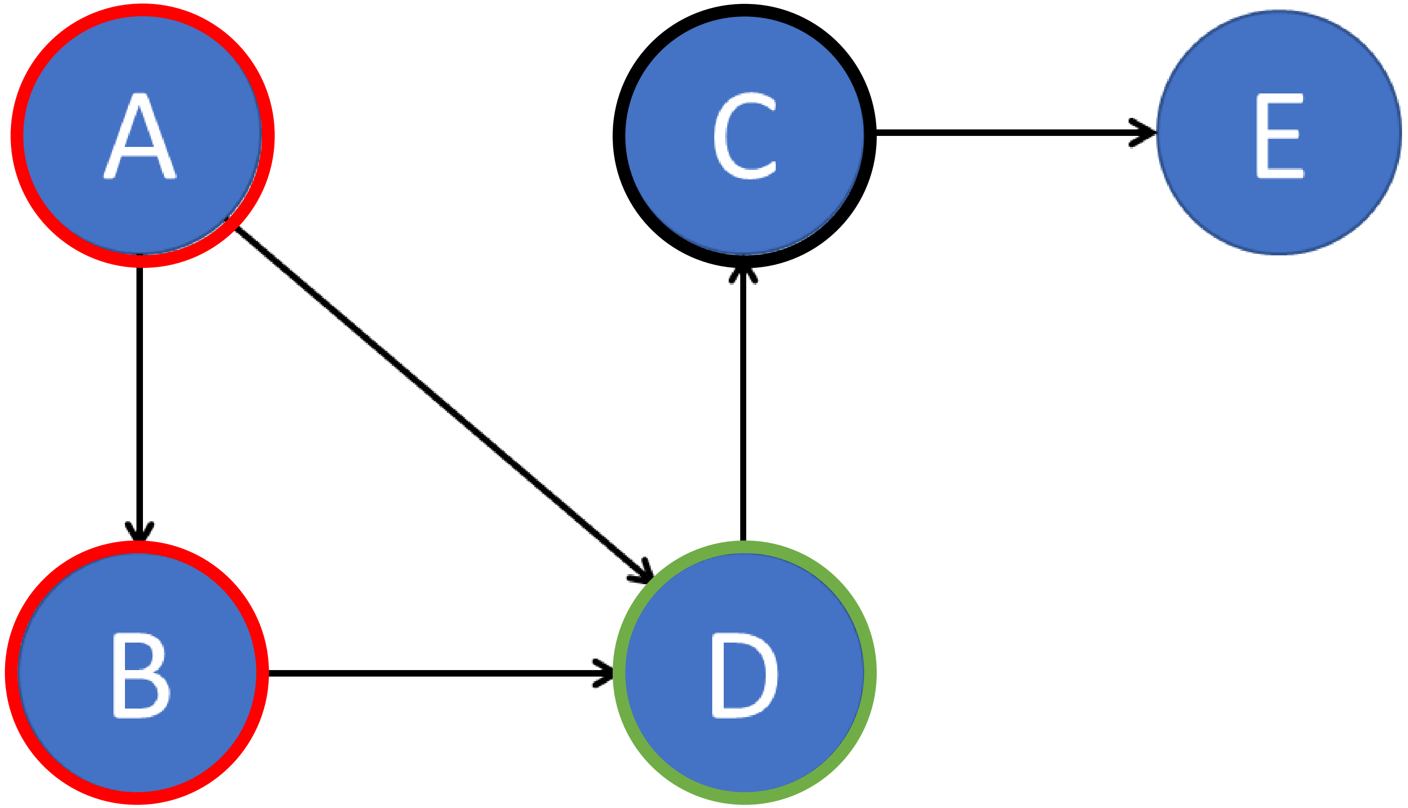
Testable Implications of Assumptions

Identification of Causal Effect

- Obtaining causal effect from observational data.

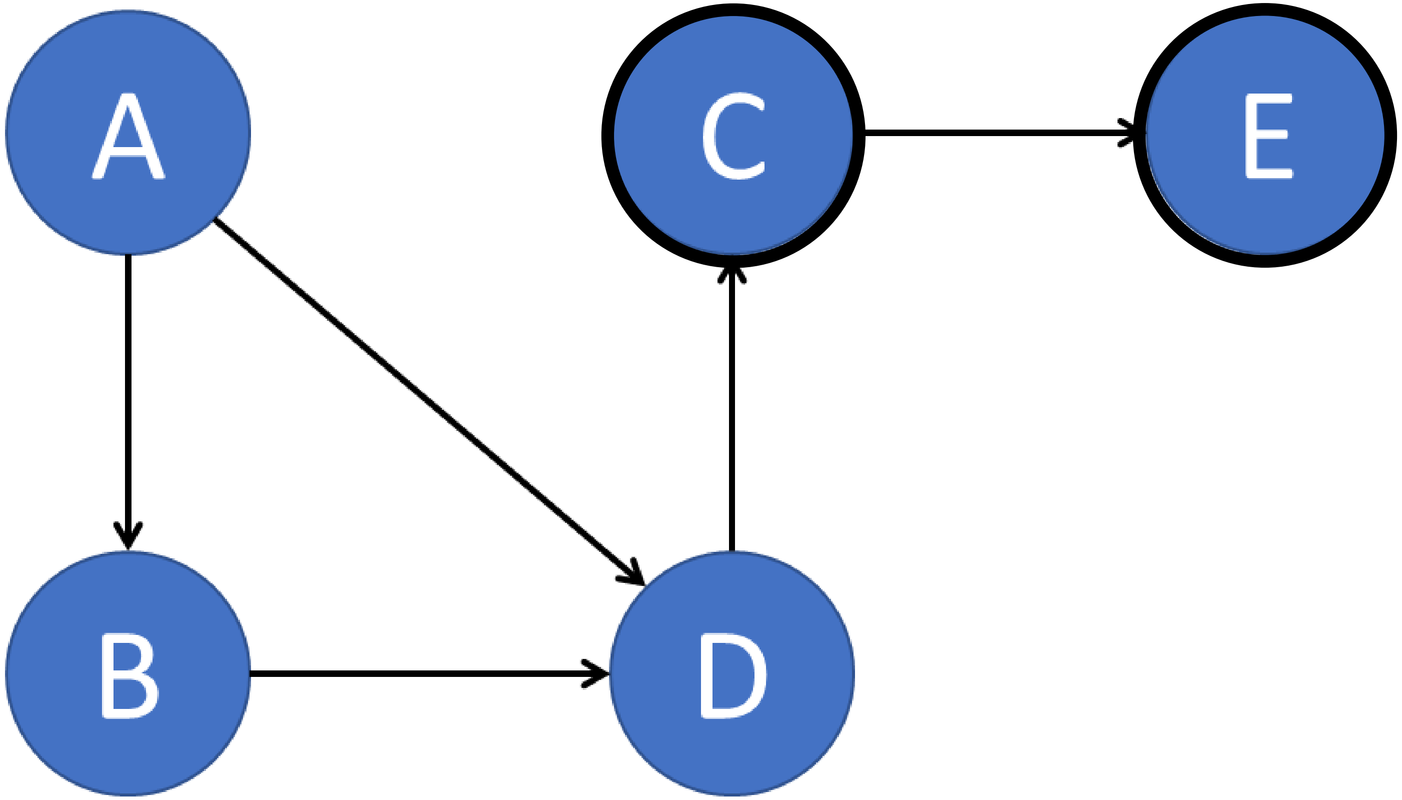
Minimality Assumption

- We only need to know the parents
  - We don't need to know A and B



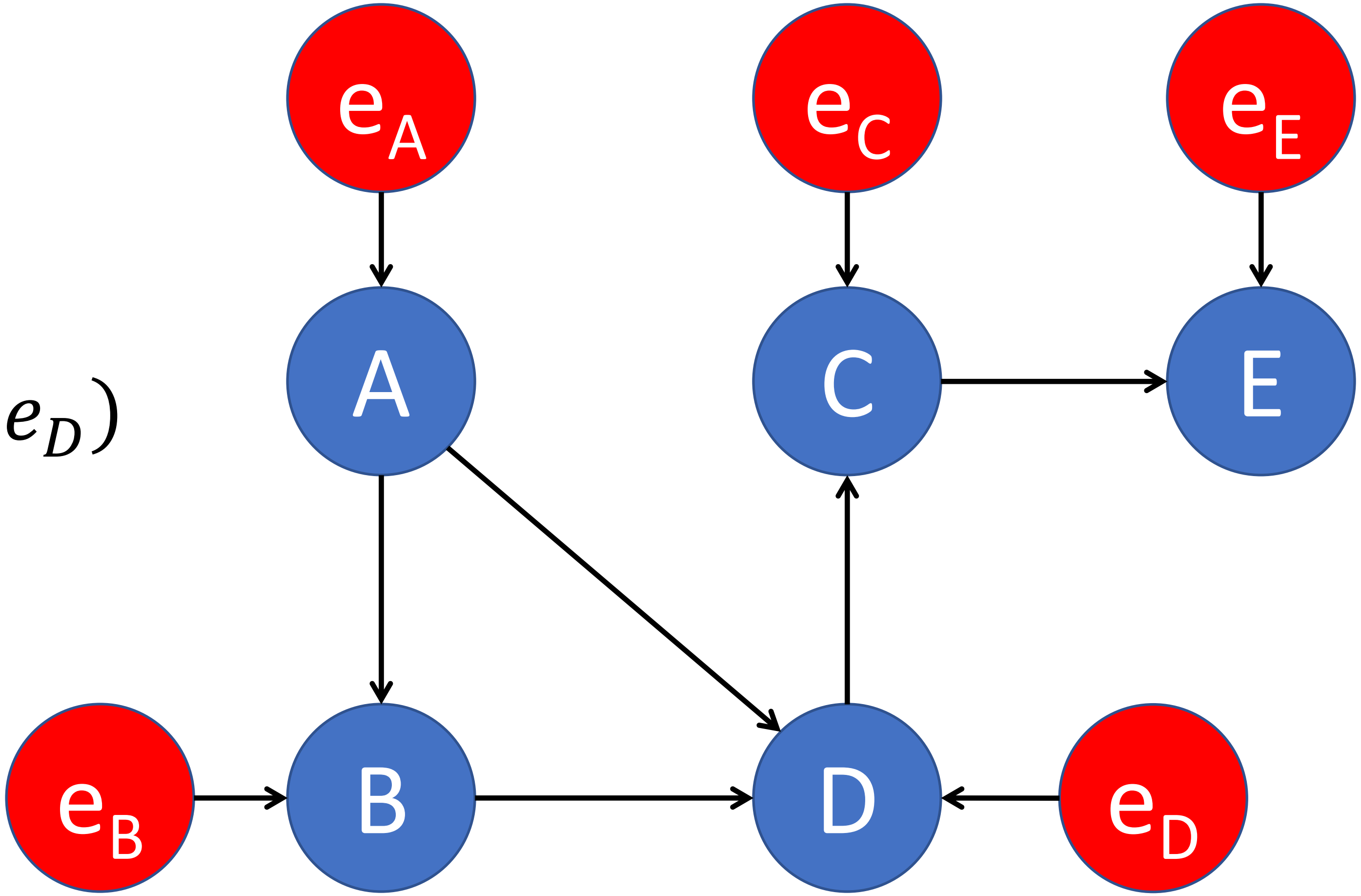
Minimality Assumption

- Adjacent nodes are dependent.
    - C and E, for example



Error Terms/Omitted Factors

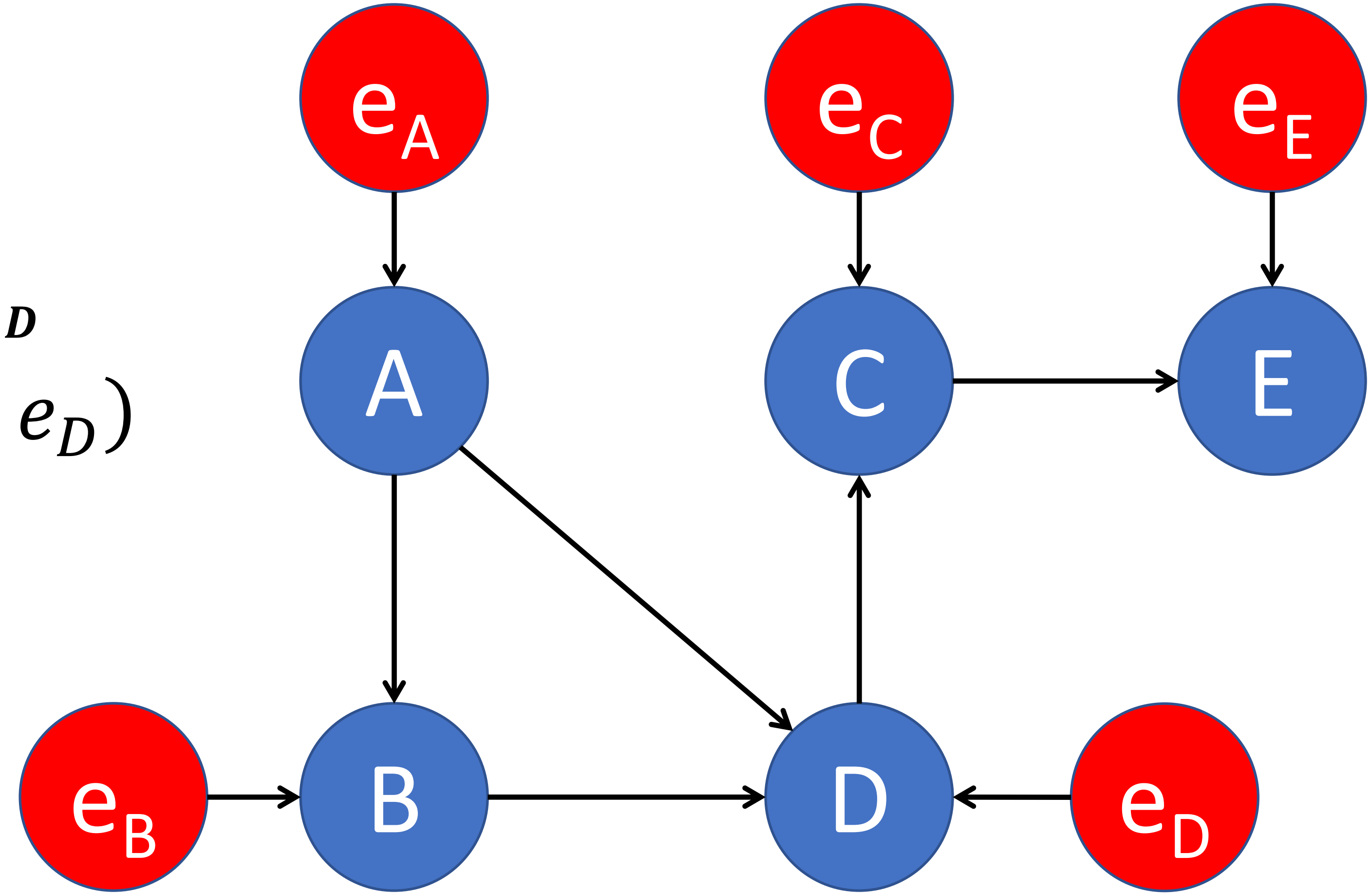
$$D = f_D(A, B, e_D)$$



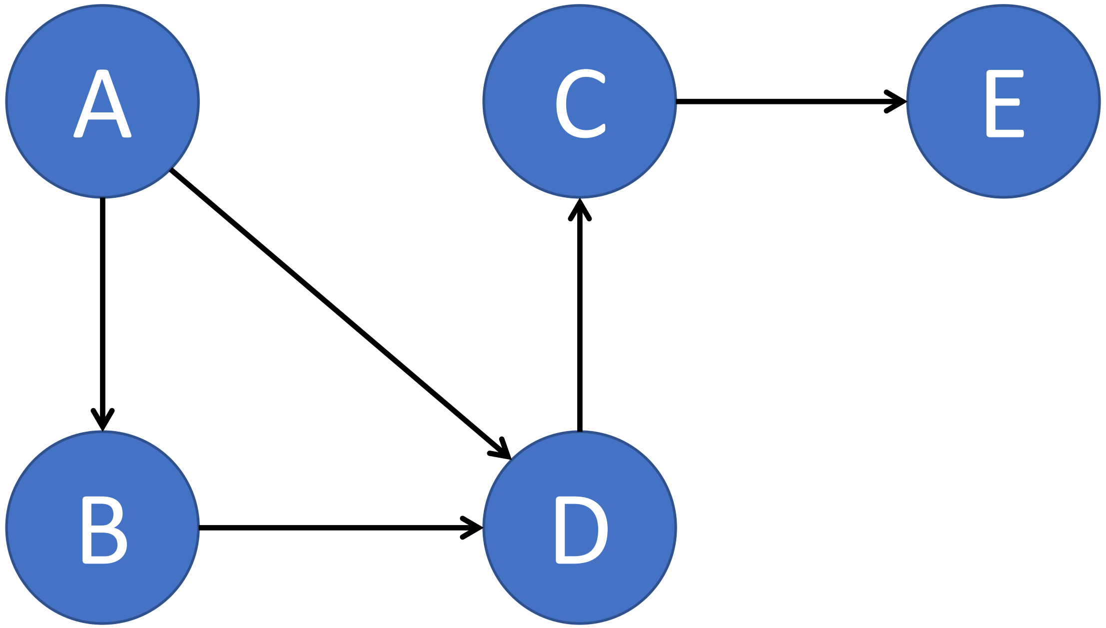
Error Terms/Omitted Factors

Parents of *D*

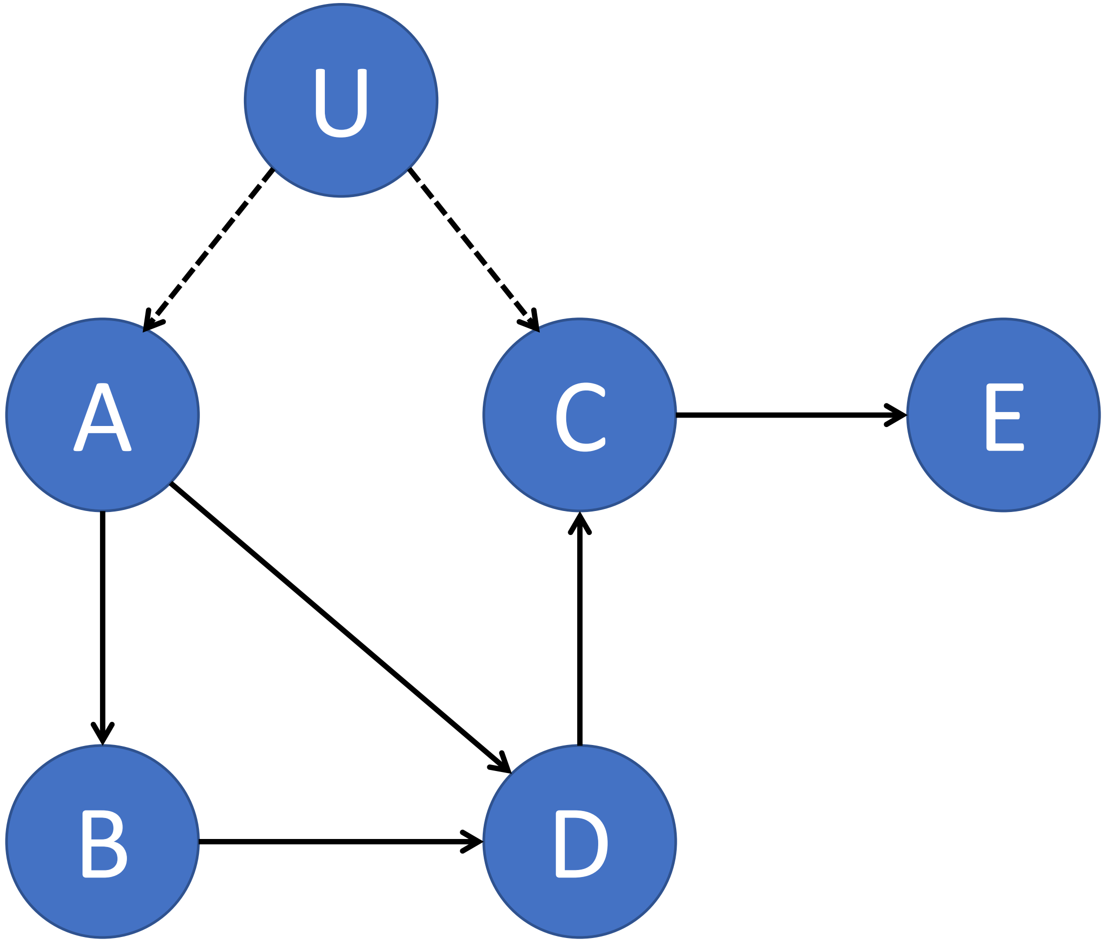
$$D = f_D(A, B, e_D)$$



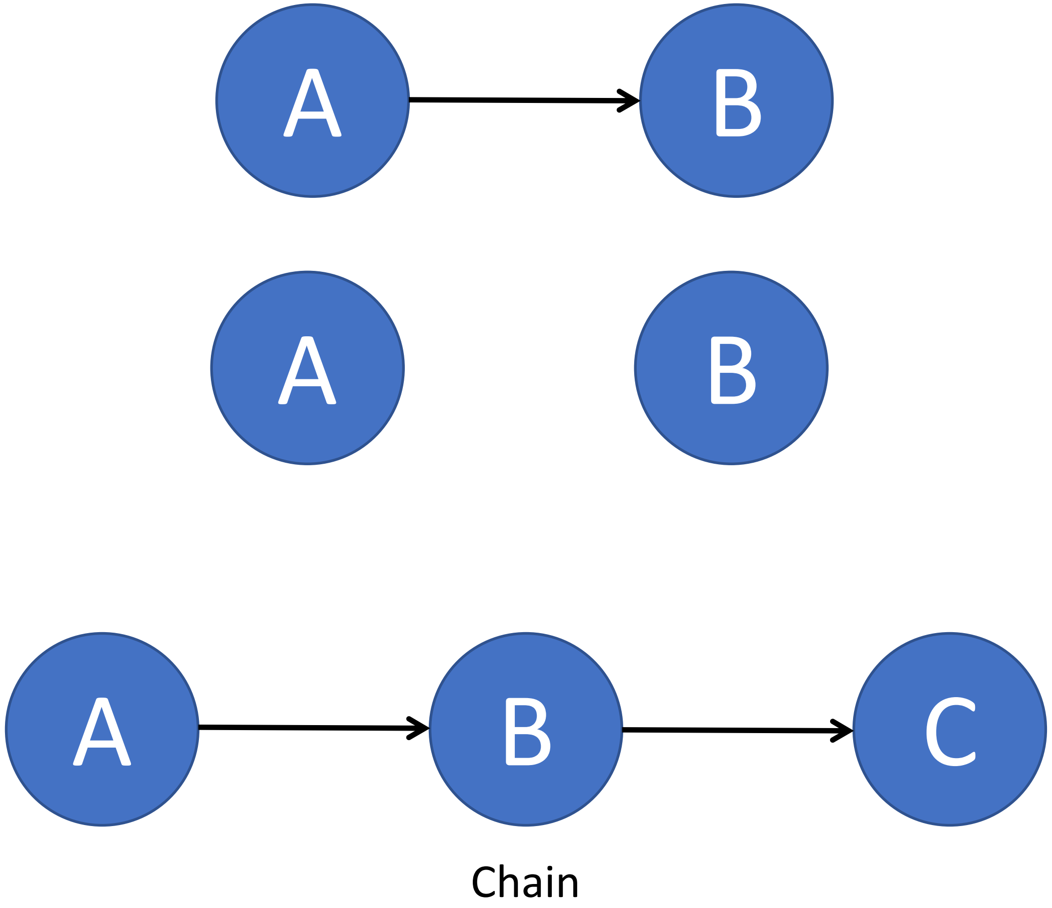
Error Terms/Omitted Factors

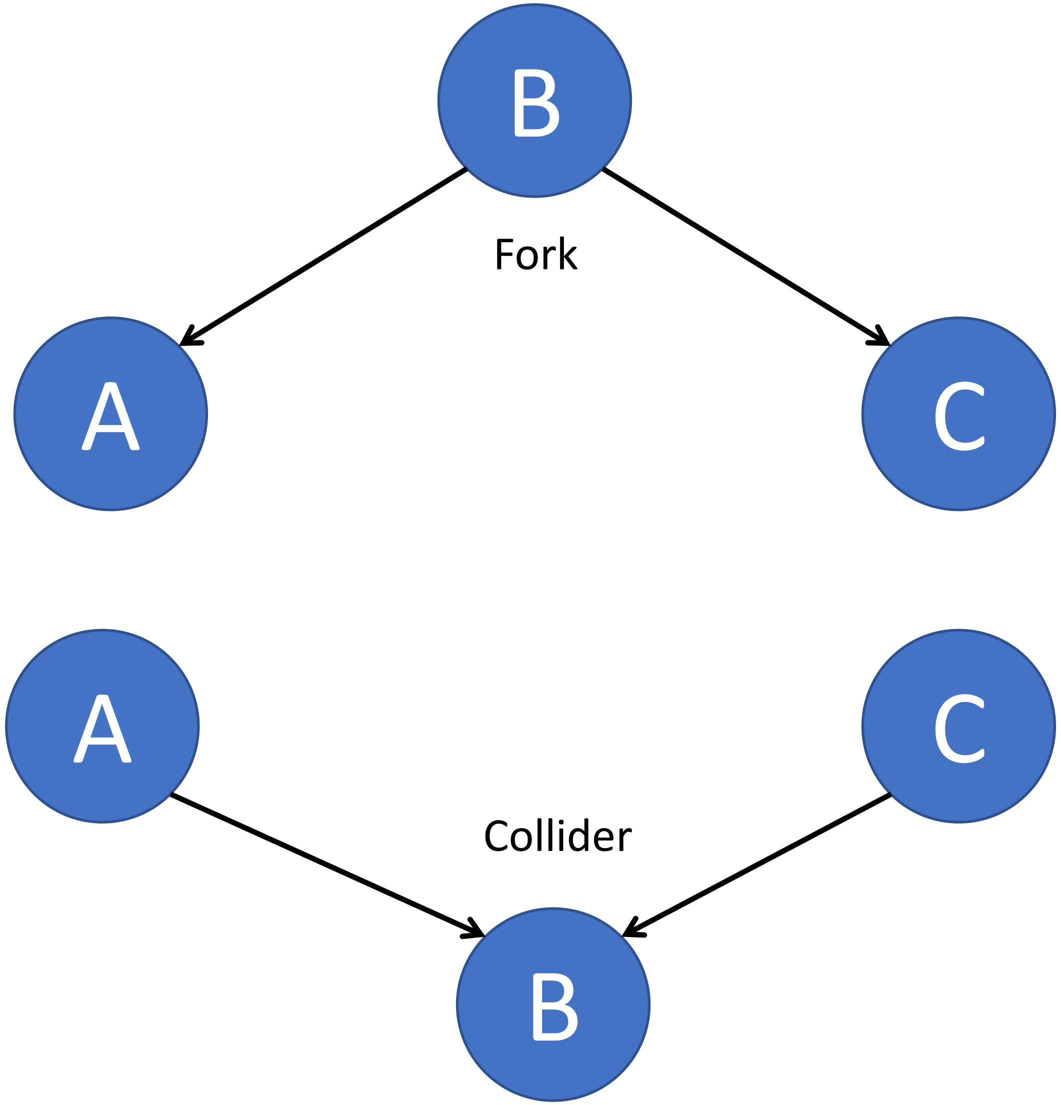


Unmeasured Variable

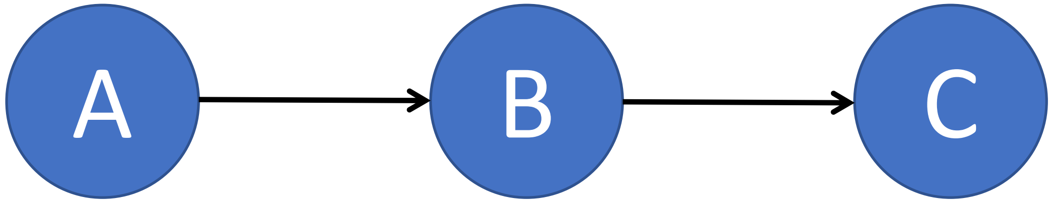


Different Configurations





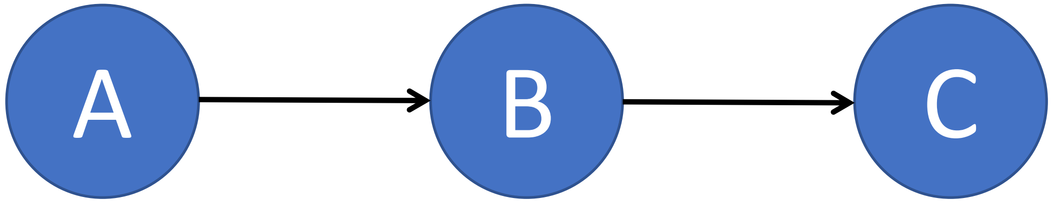
Chain



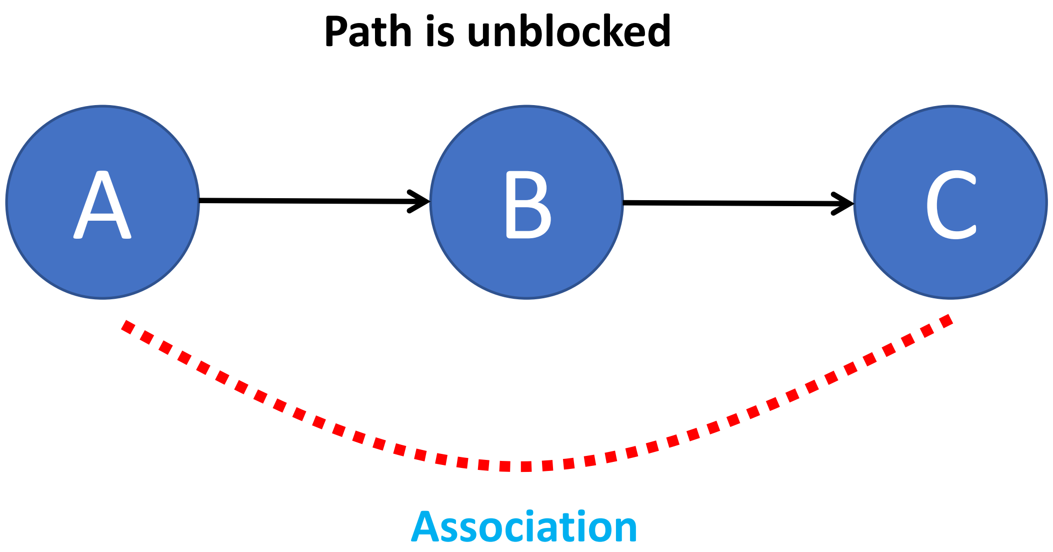
Chain

Unblocked Path ≡ Flow of Association

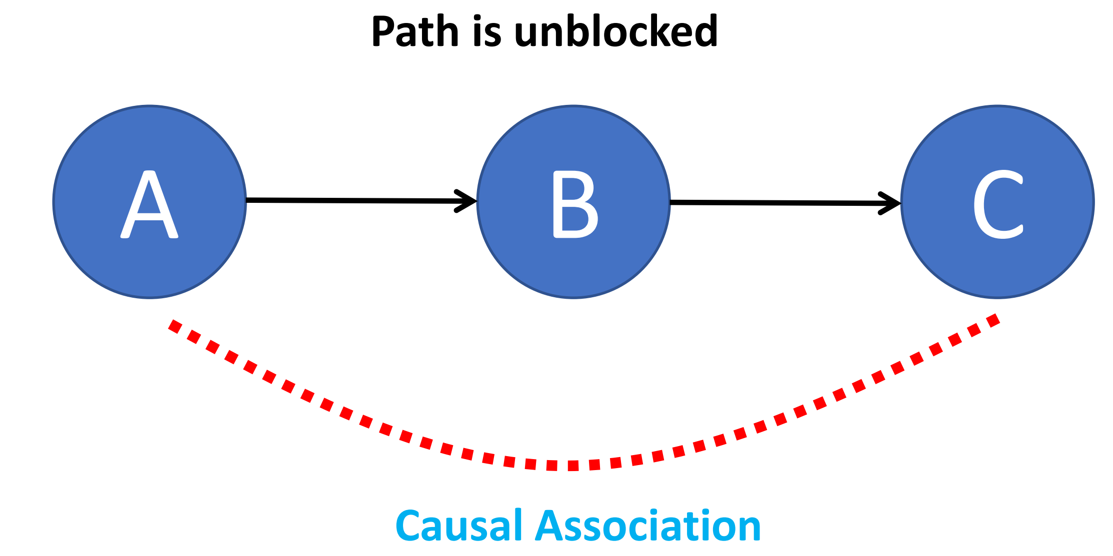
Path is unblocked



Chain

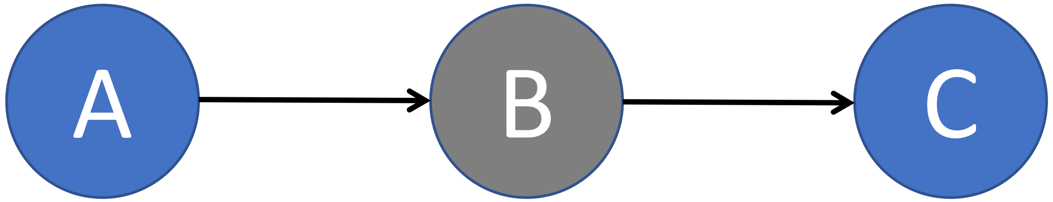


Chain



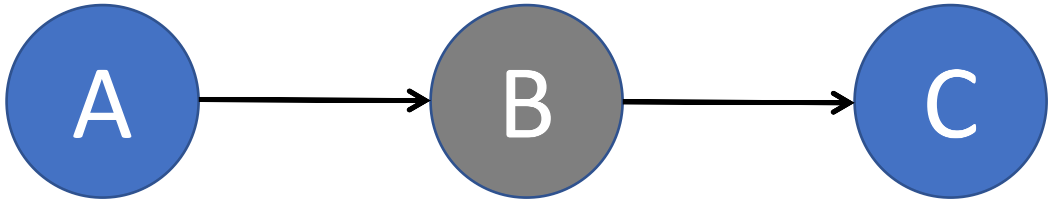
Chain

Conditioning on B

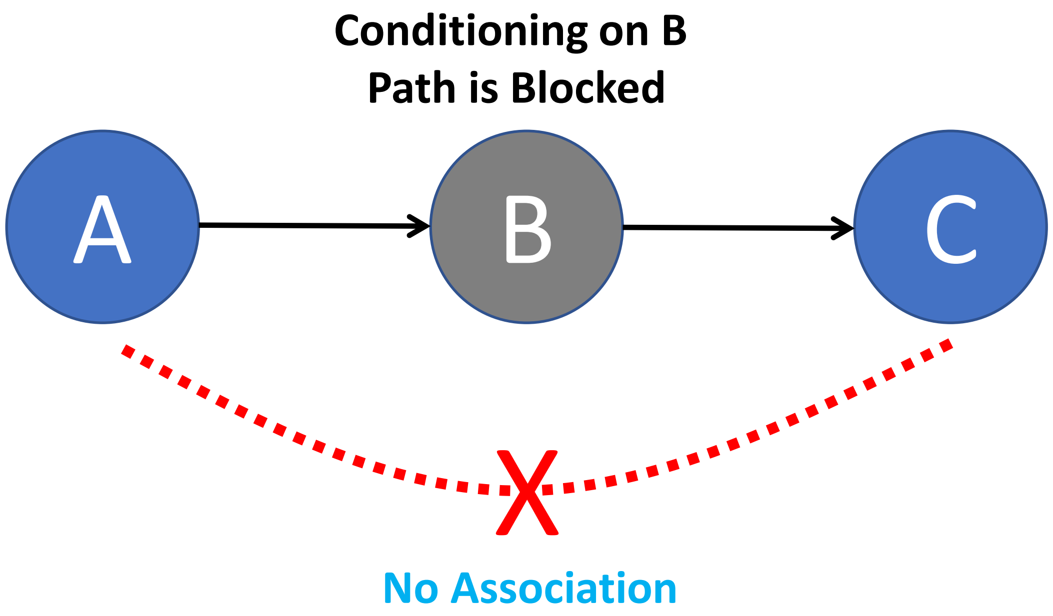


Chain

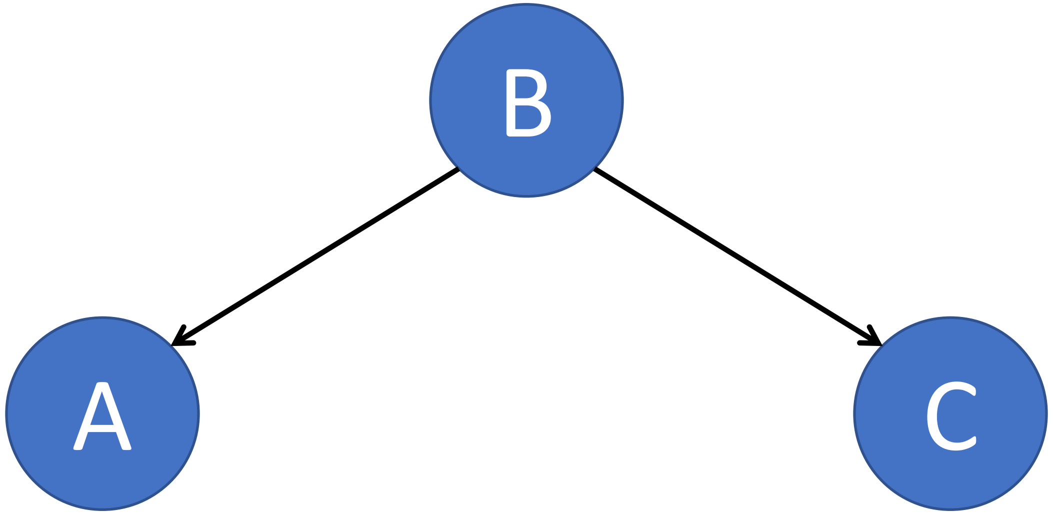
Conditioning on B
Path is Blocked



Chain

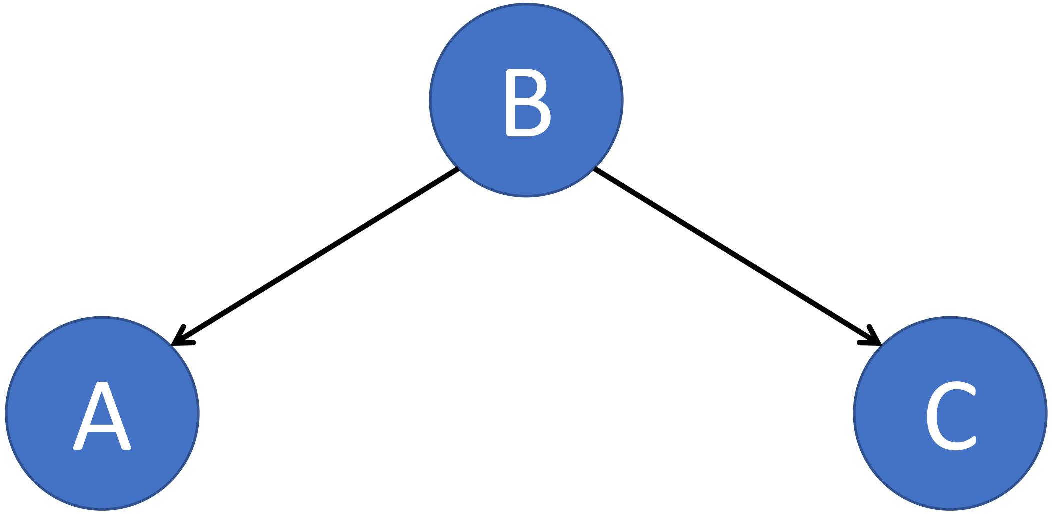


Fork

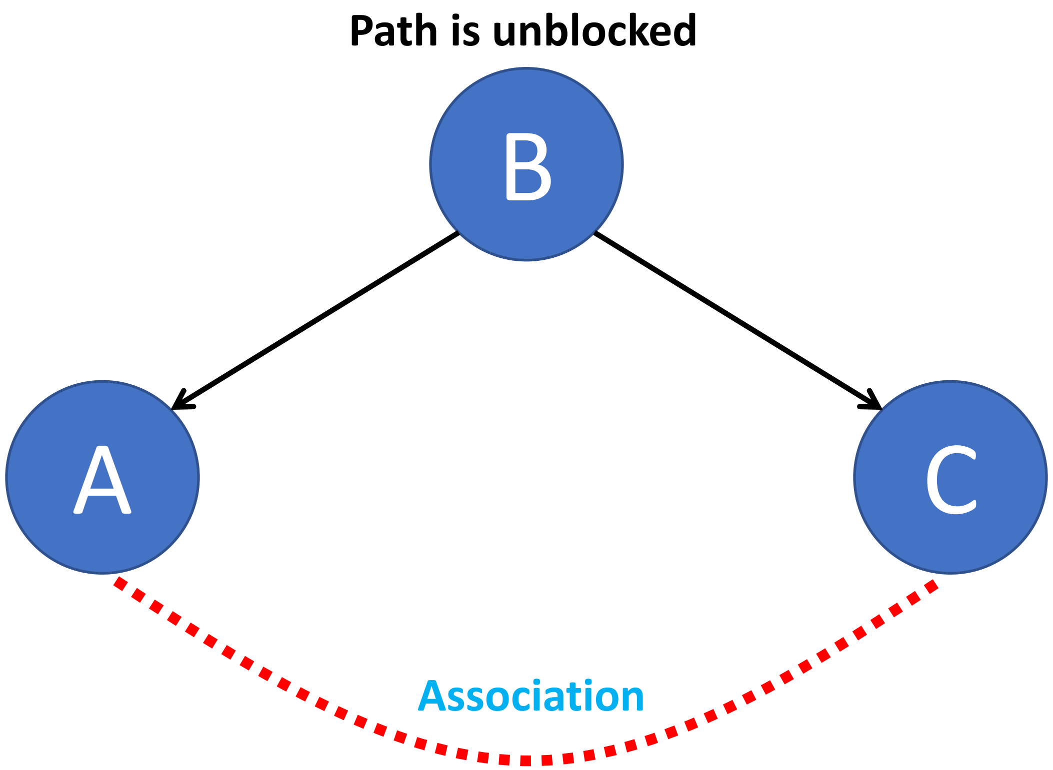


Fork

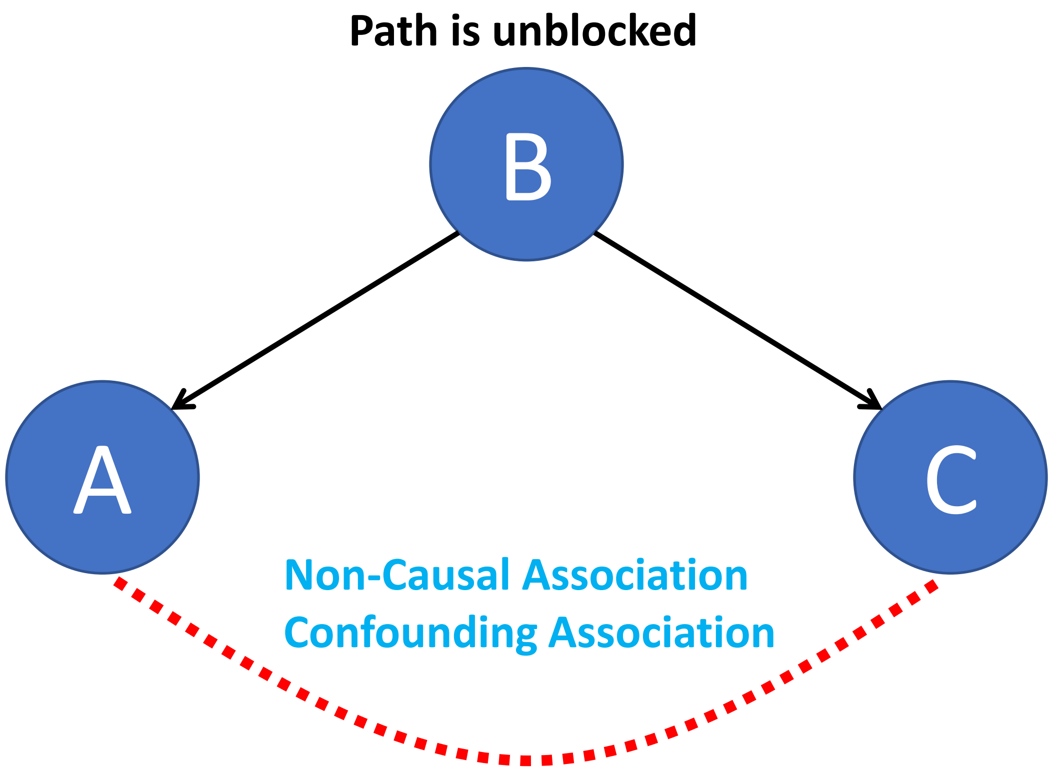
Path is unblocked



Fork

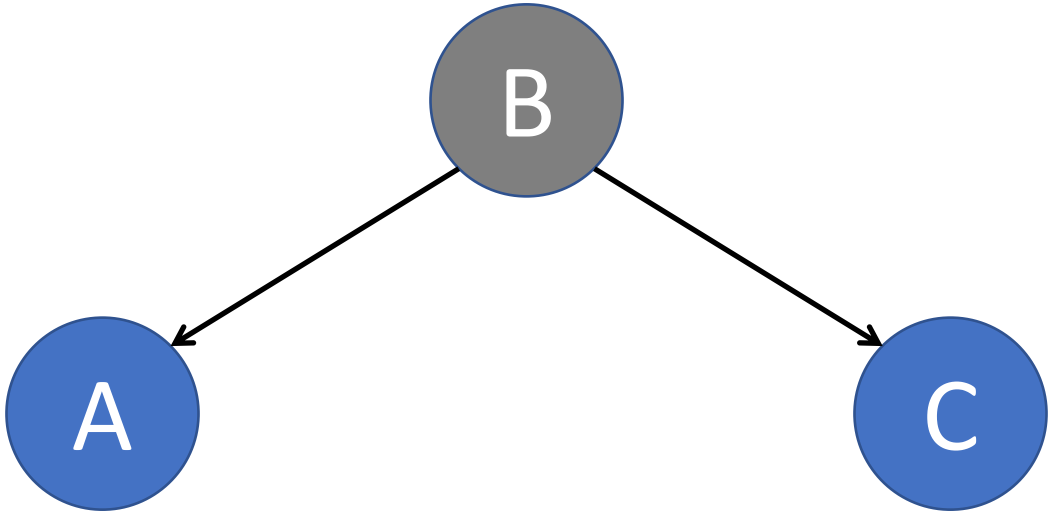


Fork

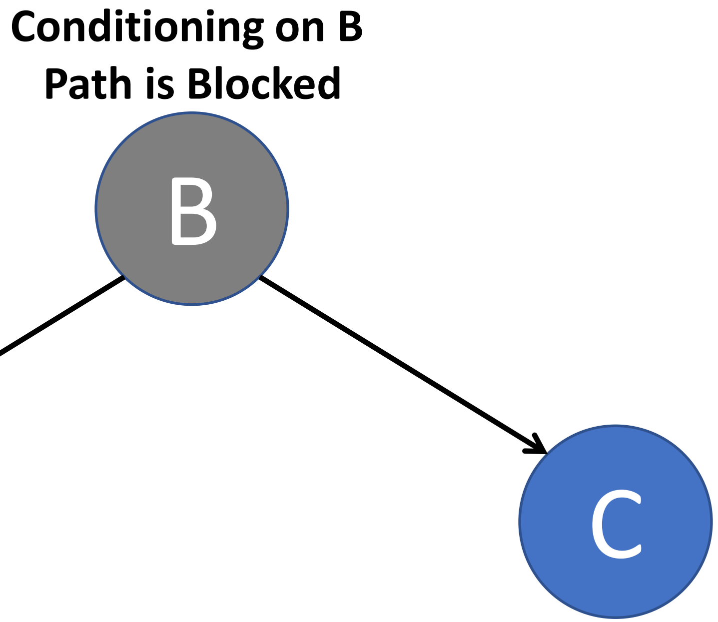

Path is unblocked
B
A
C
Non-Causal Association
Confounding Association

Fork

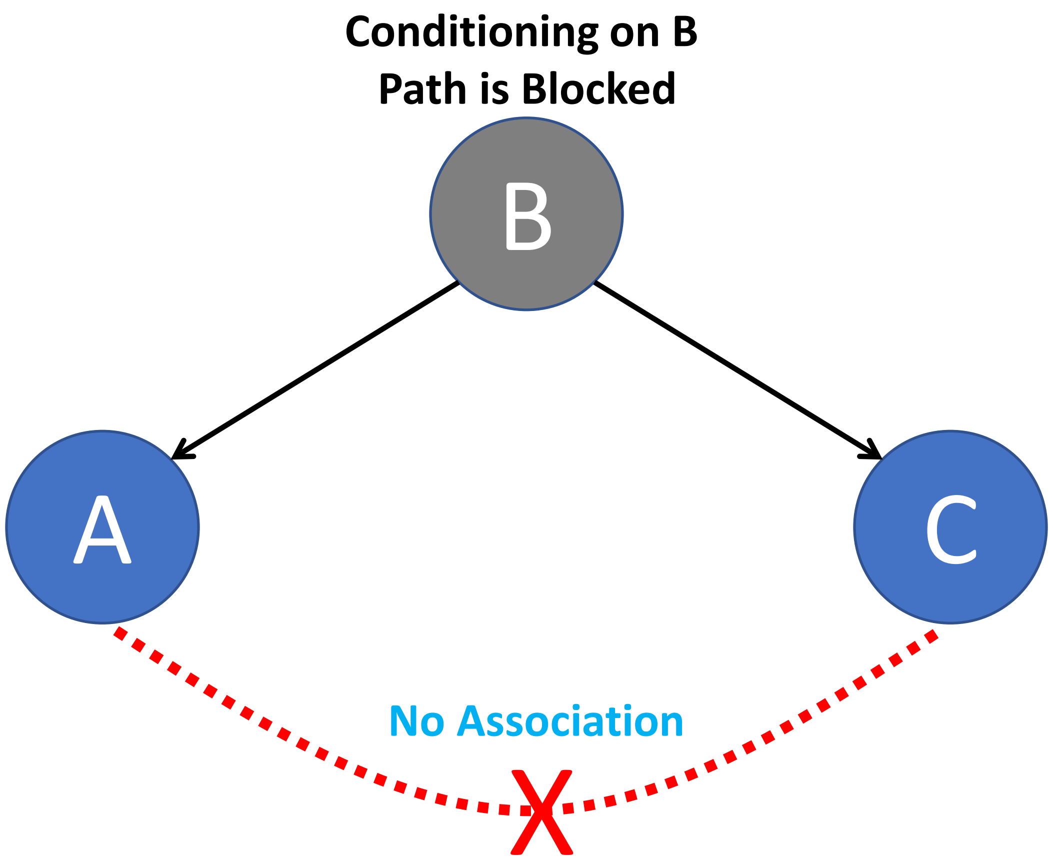
Conditioning on B



Fork

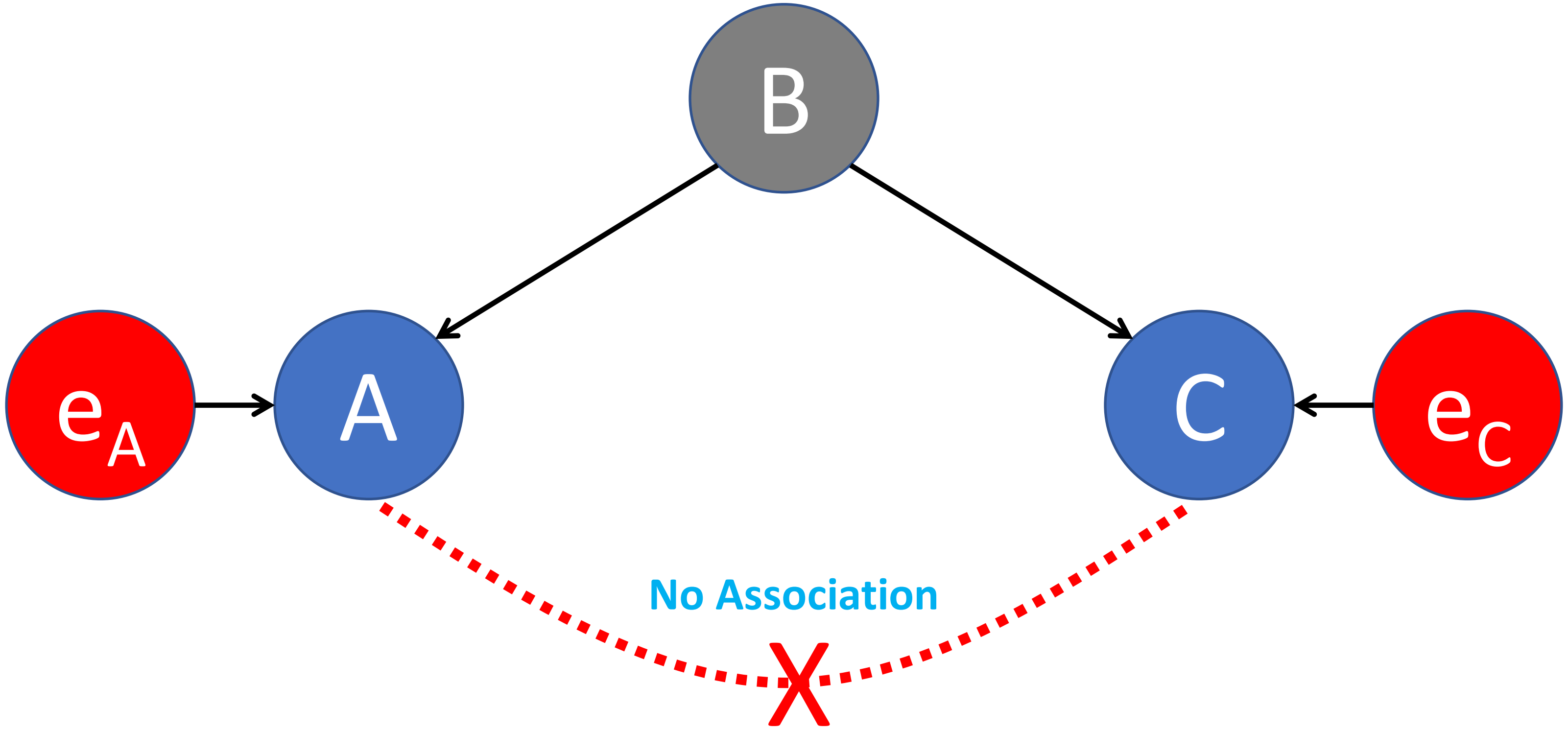


Fork

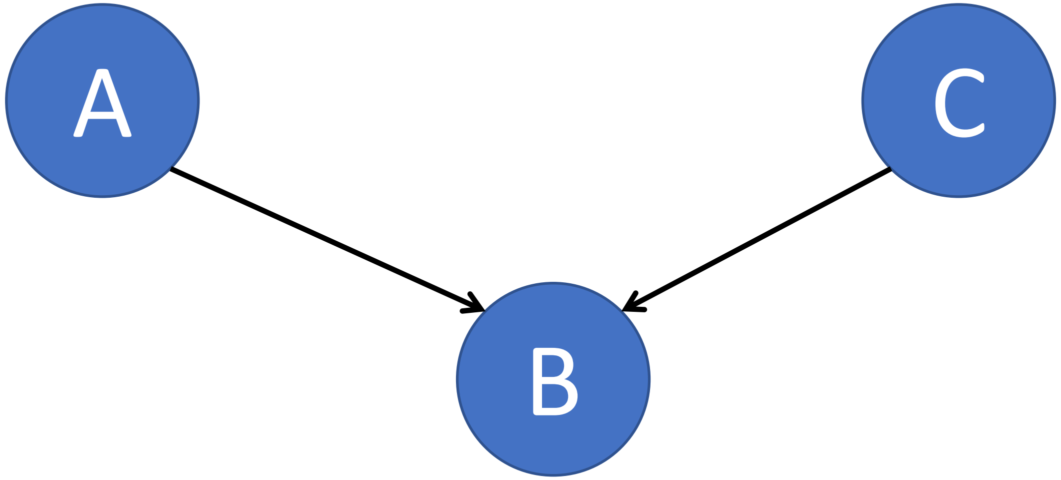


Fork

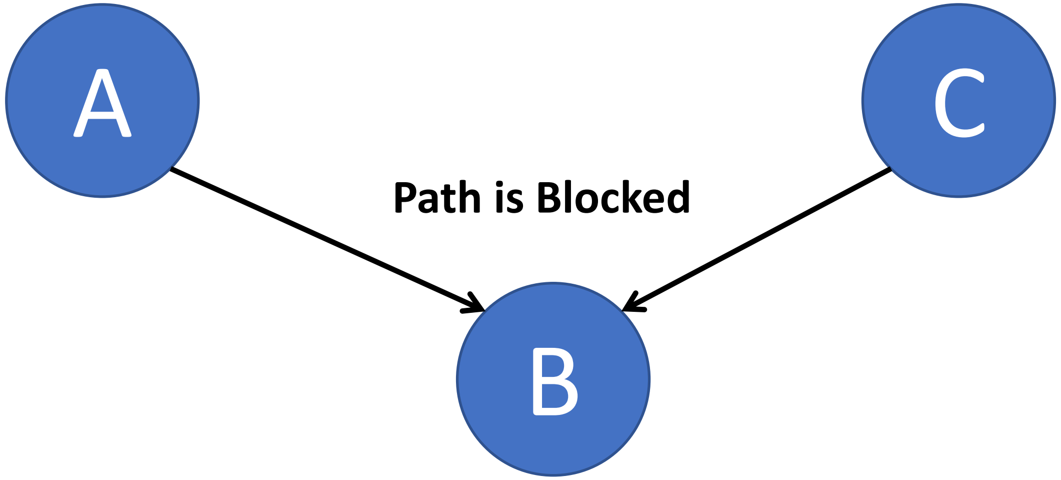
Conditioning on B
Path is Blocked



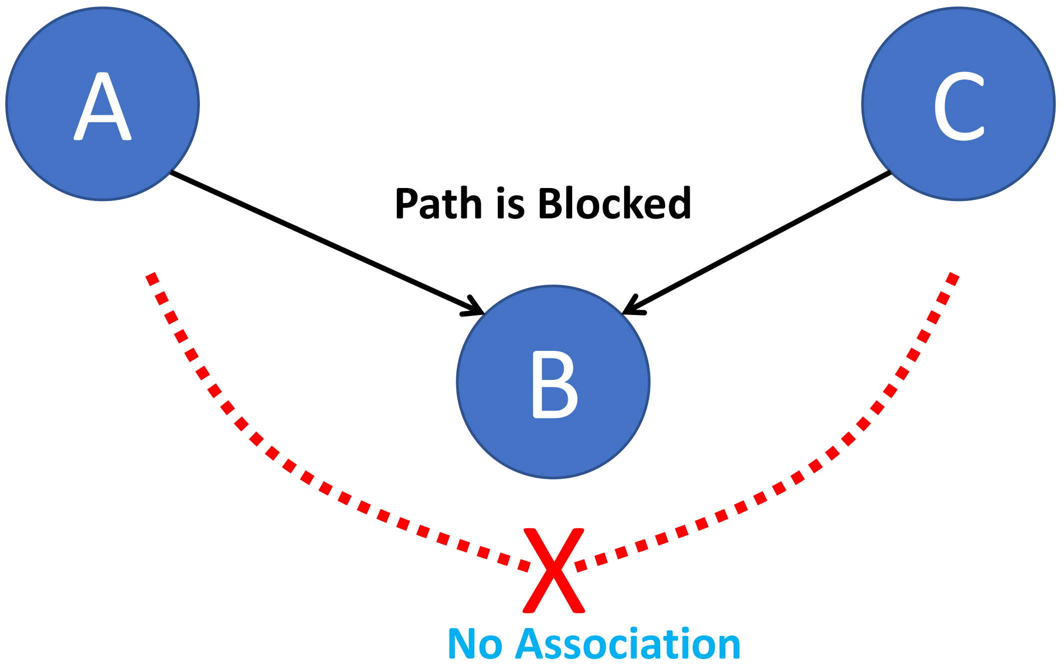
Collider



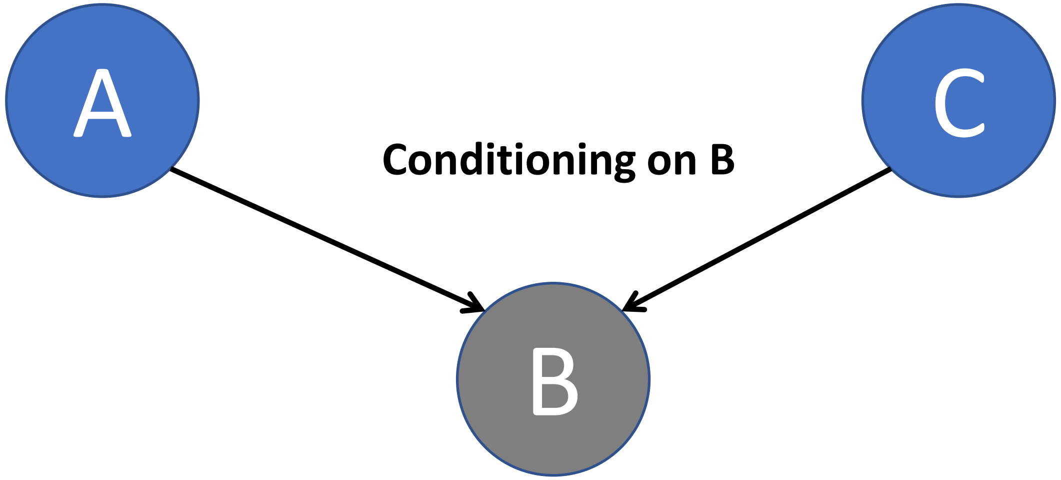
Collider



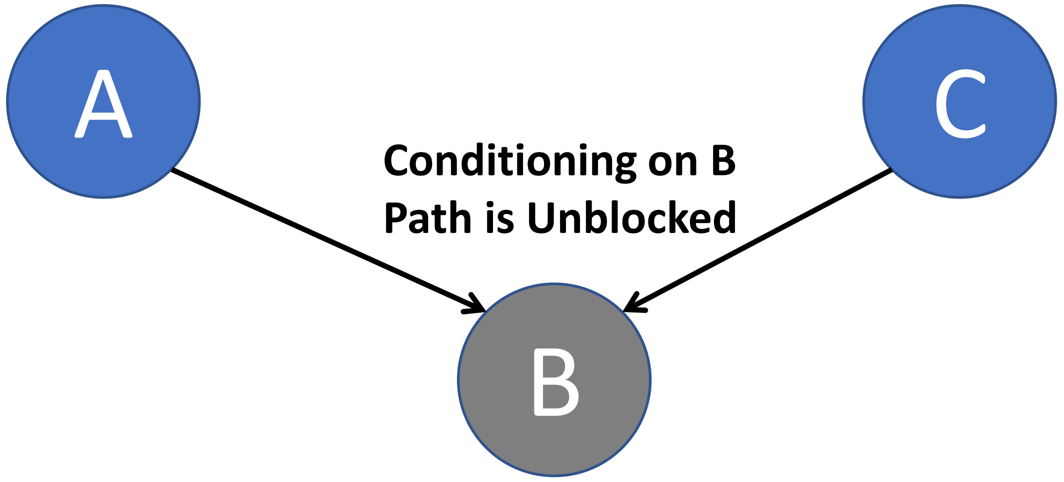
Collider



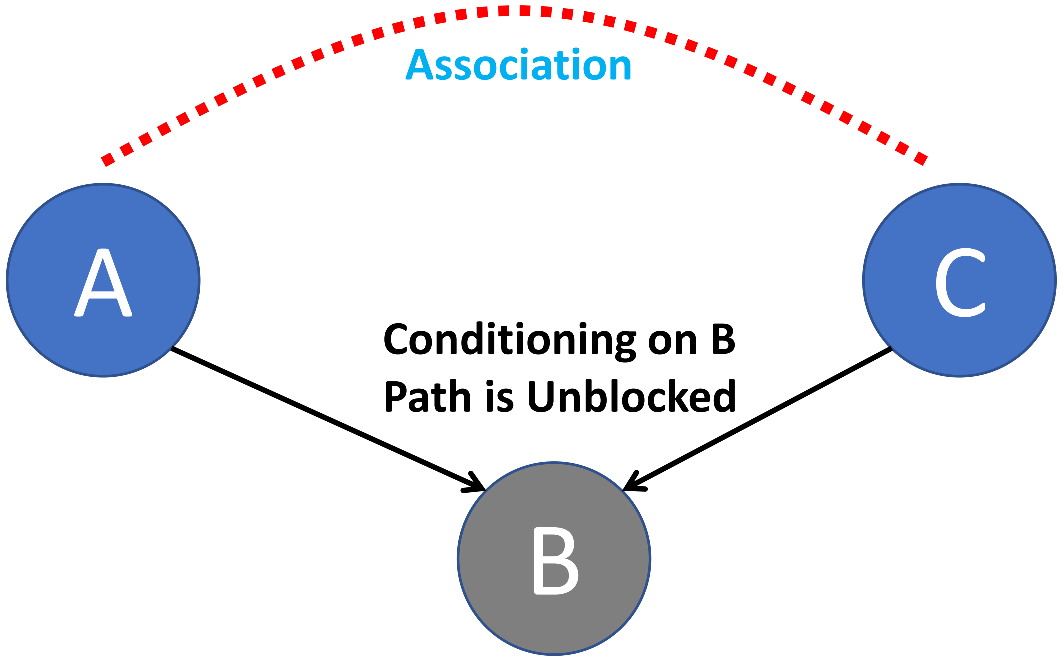
Collider



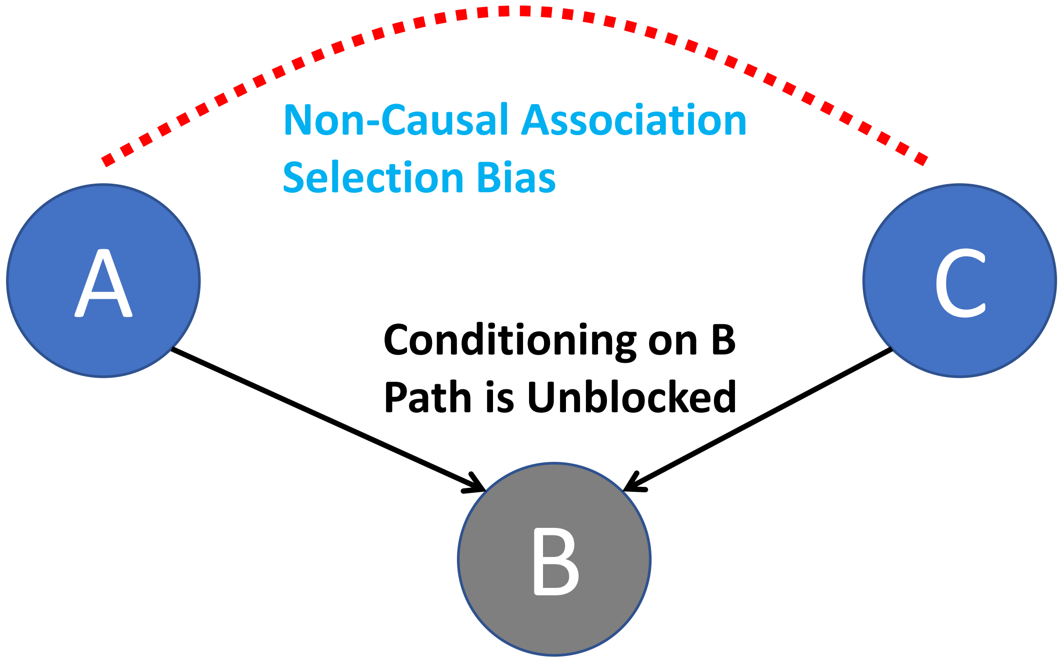
Collider



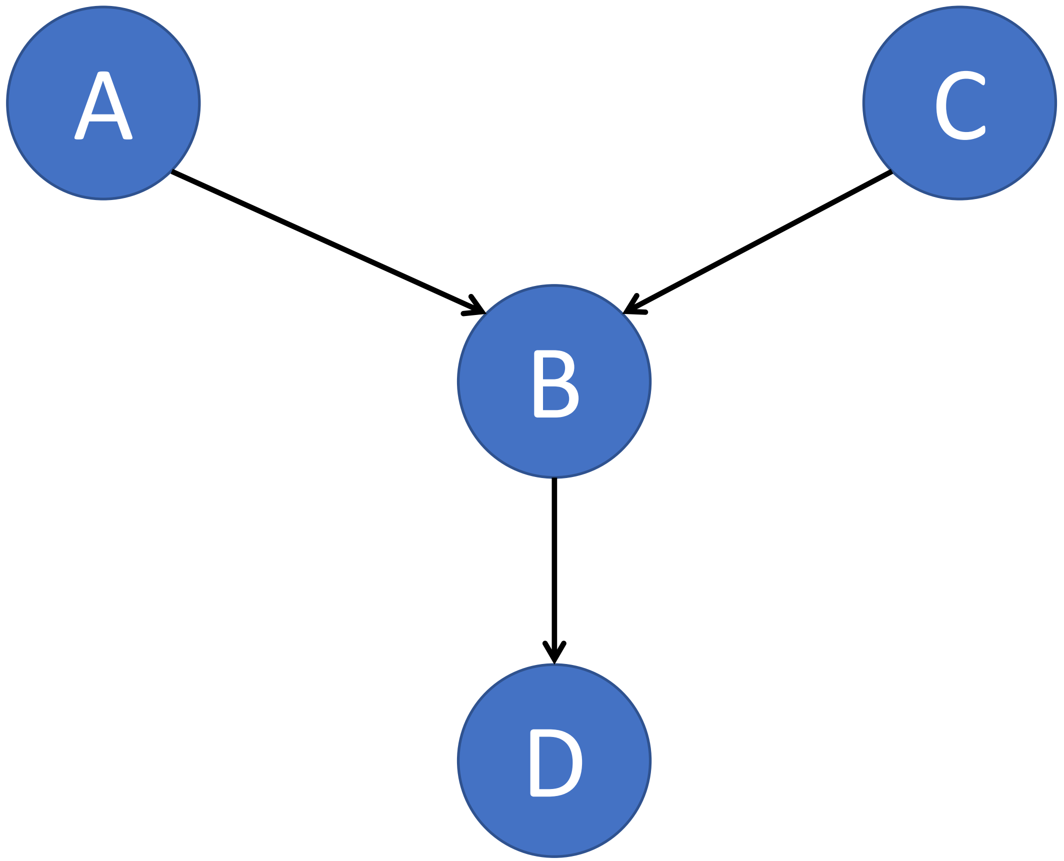
Collider



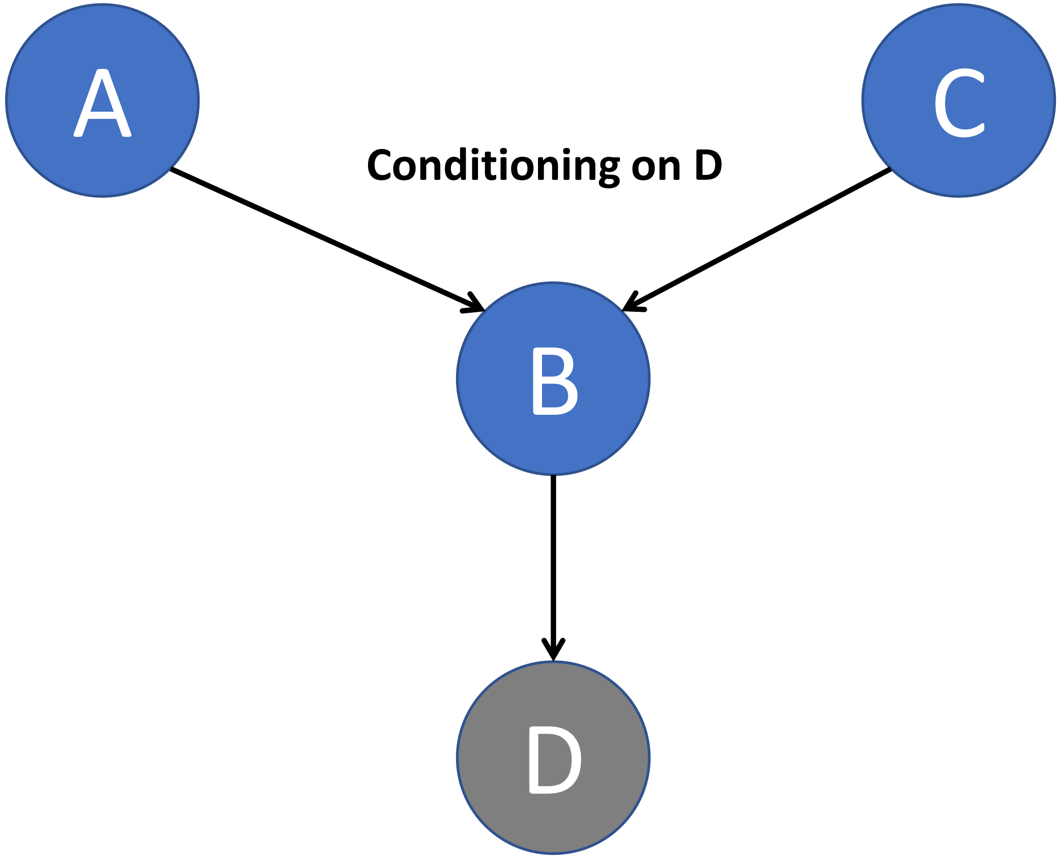
Collider


Non-Causal Association
Selection Bias
A
C
Conditioning on B
Path is Unblocked
B

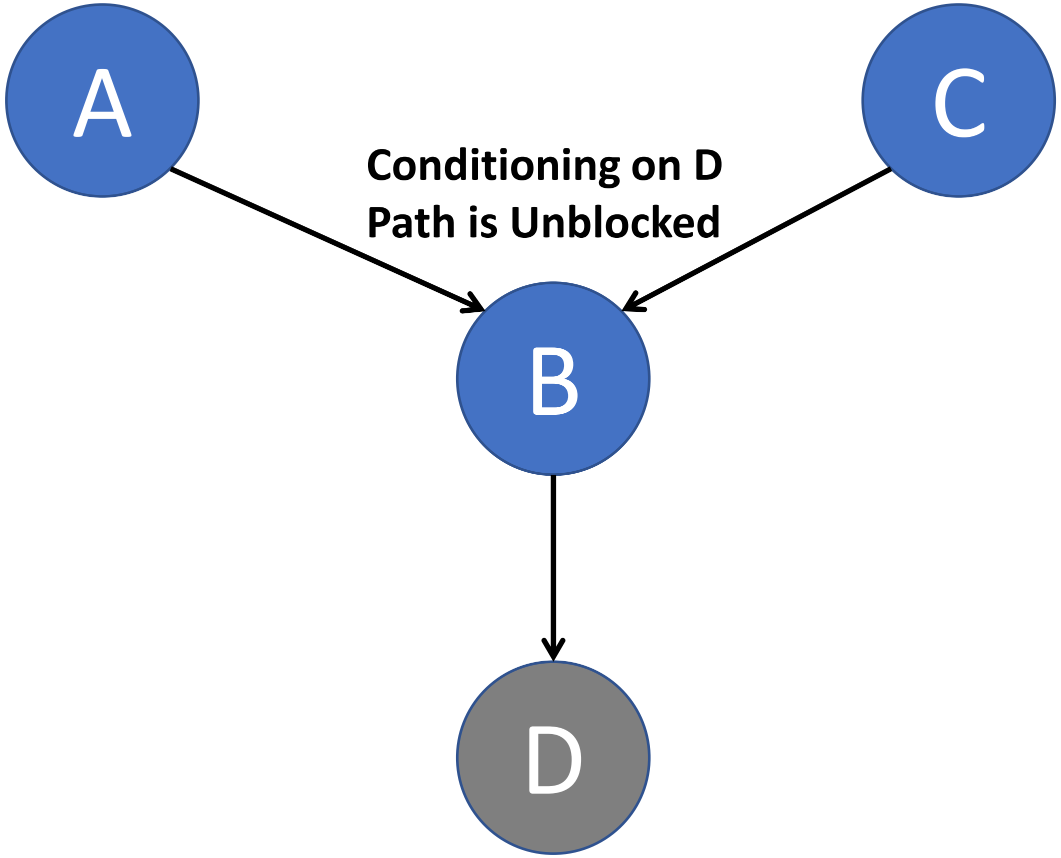
Collider



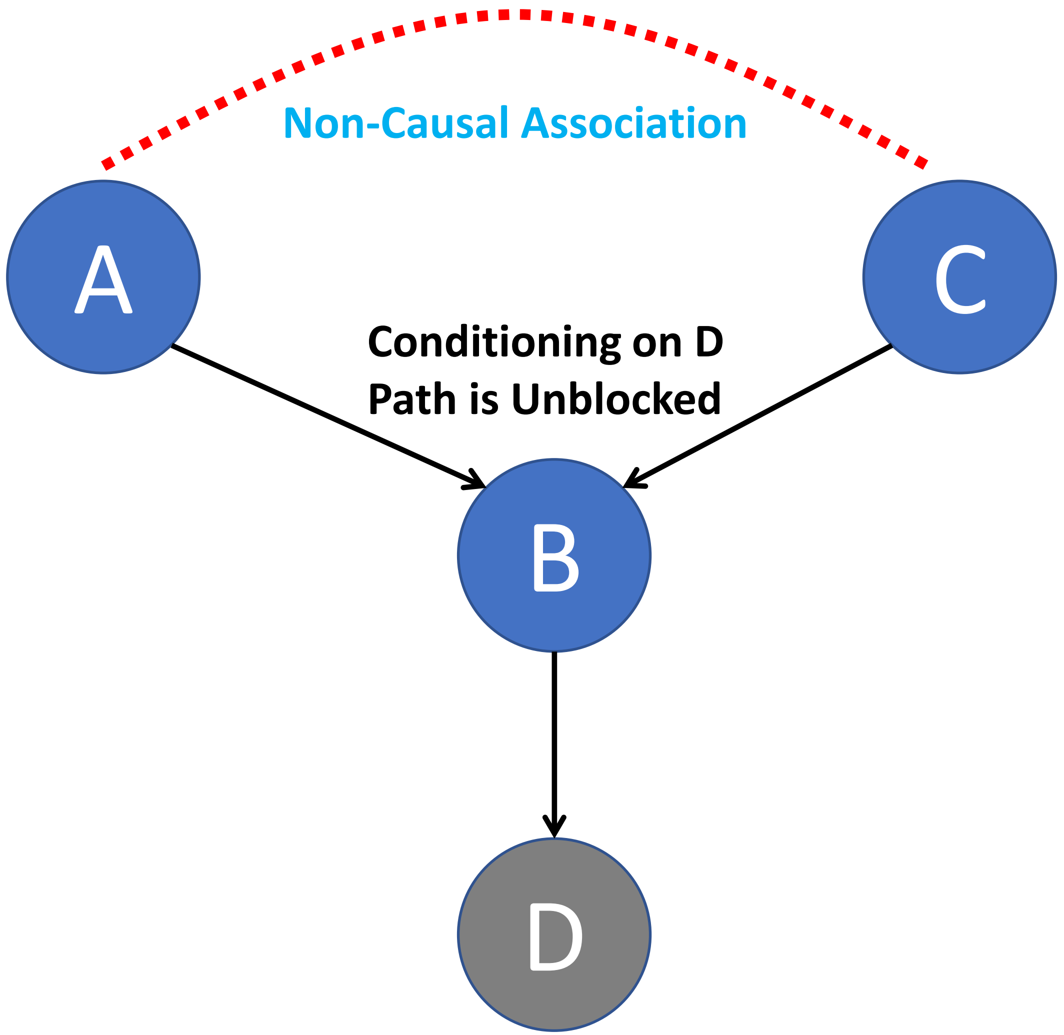
Collider



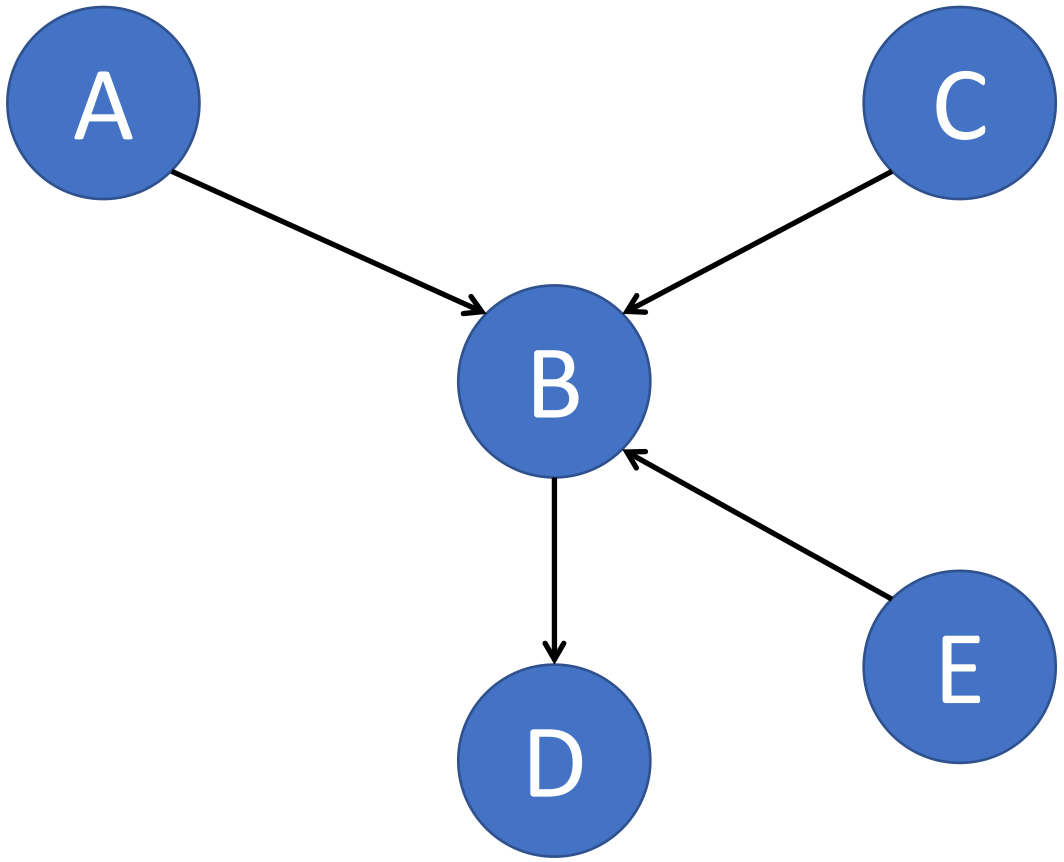
Collider



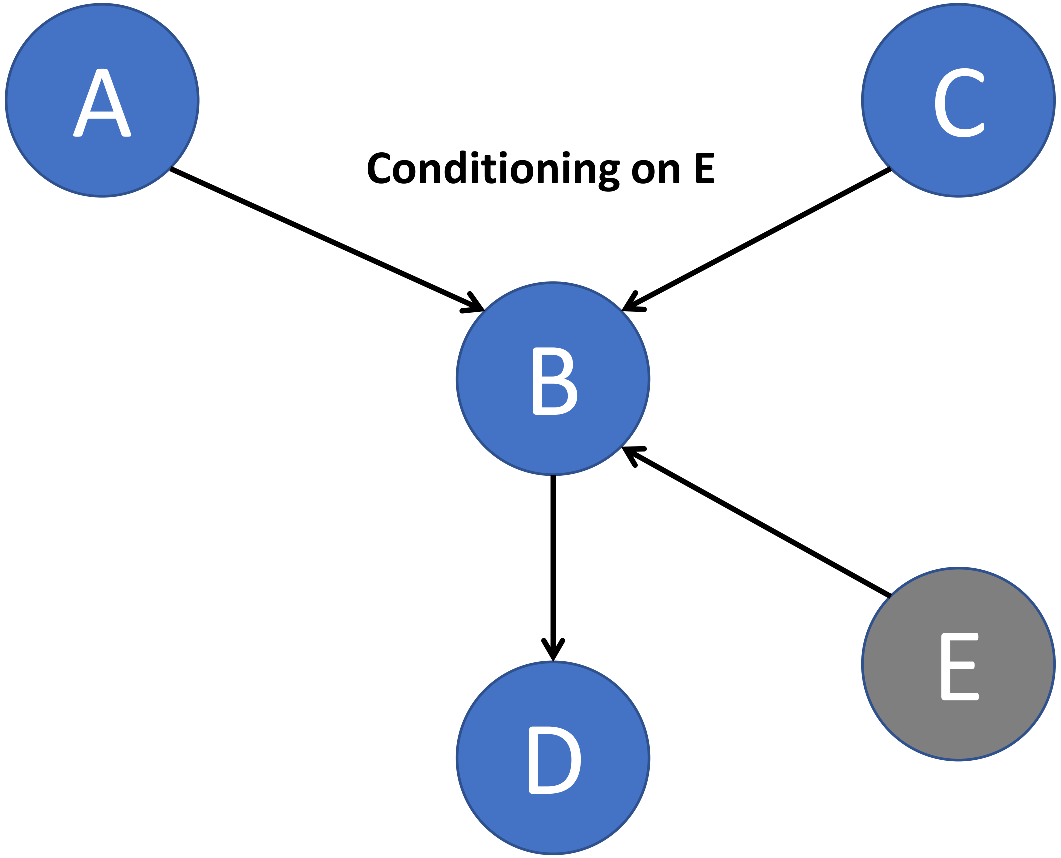
Collider


Non-Causal Association
A
C
Conditioning on D
Path is Unblocked
B
D

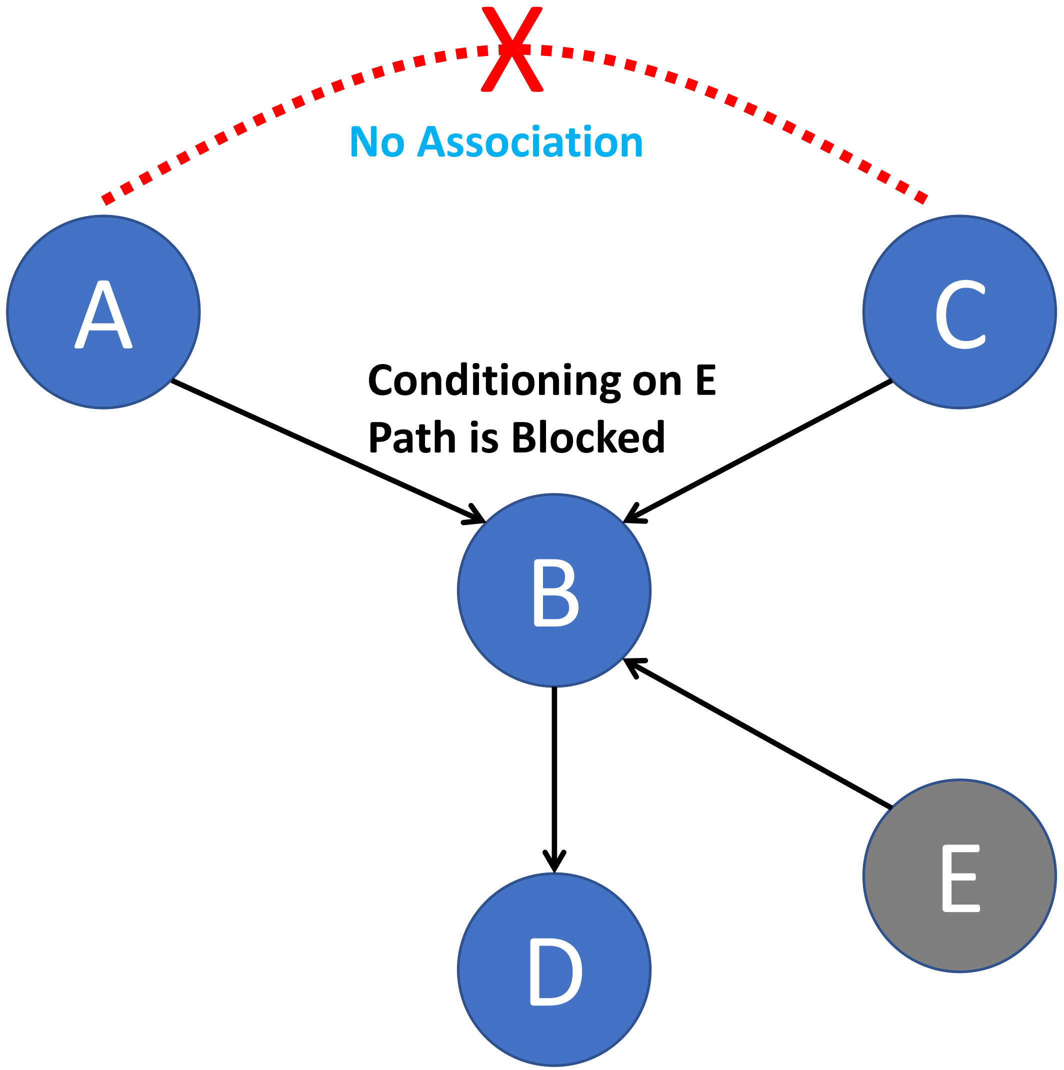
Collider



Collider



Collider



Blocked Path

- Conditioning on a set **Z** blocks a path between A and B:
  - When there is a 

 or 

 in the path and W is in **Z**
  - If there is a collider; and collider or its descendants are not in **Z**.

d-separation

Two variables A and B are d-separated by variables in **Z**, if all paths between them are blocked by **Z**.

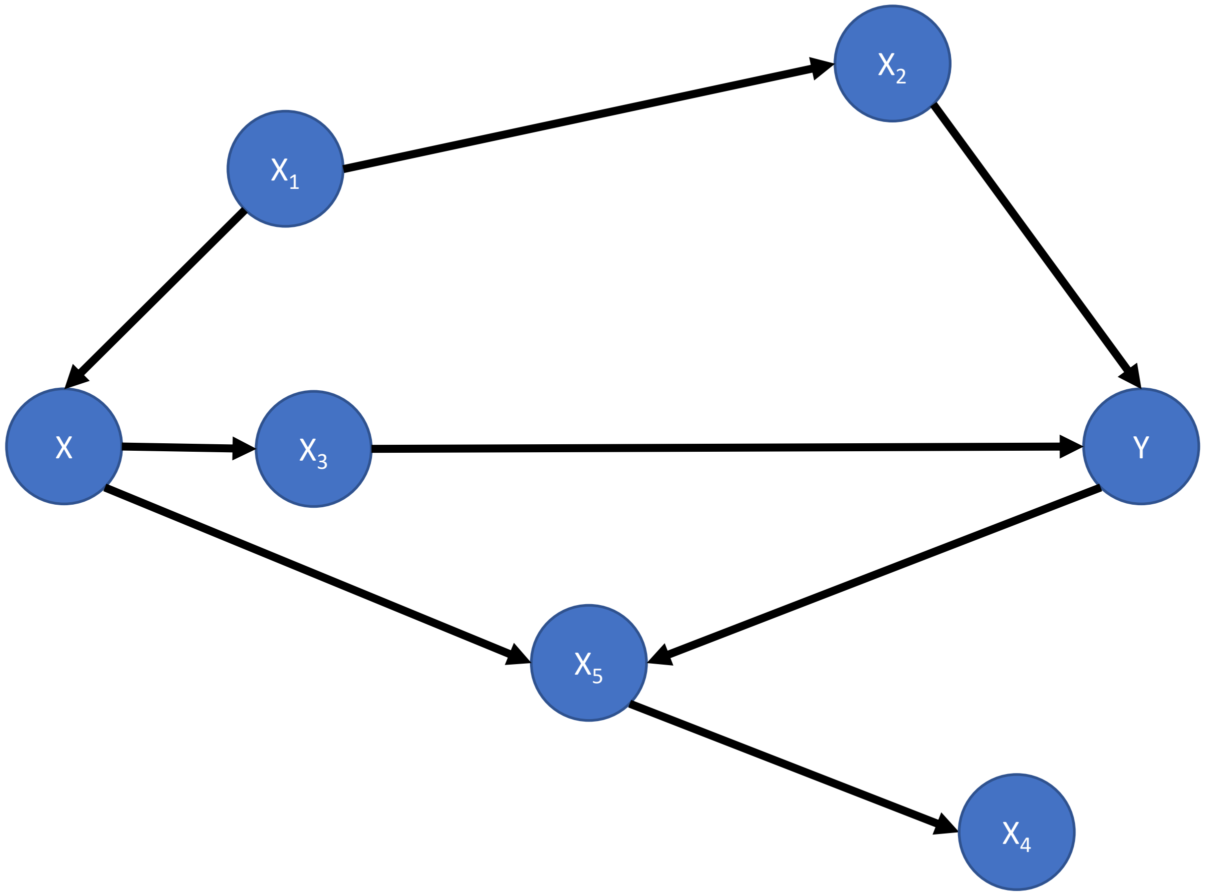
Two variables are d-connected if and only if they are not d-separated.

When A and B are d-separated by **Z**, A and B are independent conditional to **Z**.

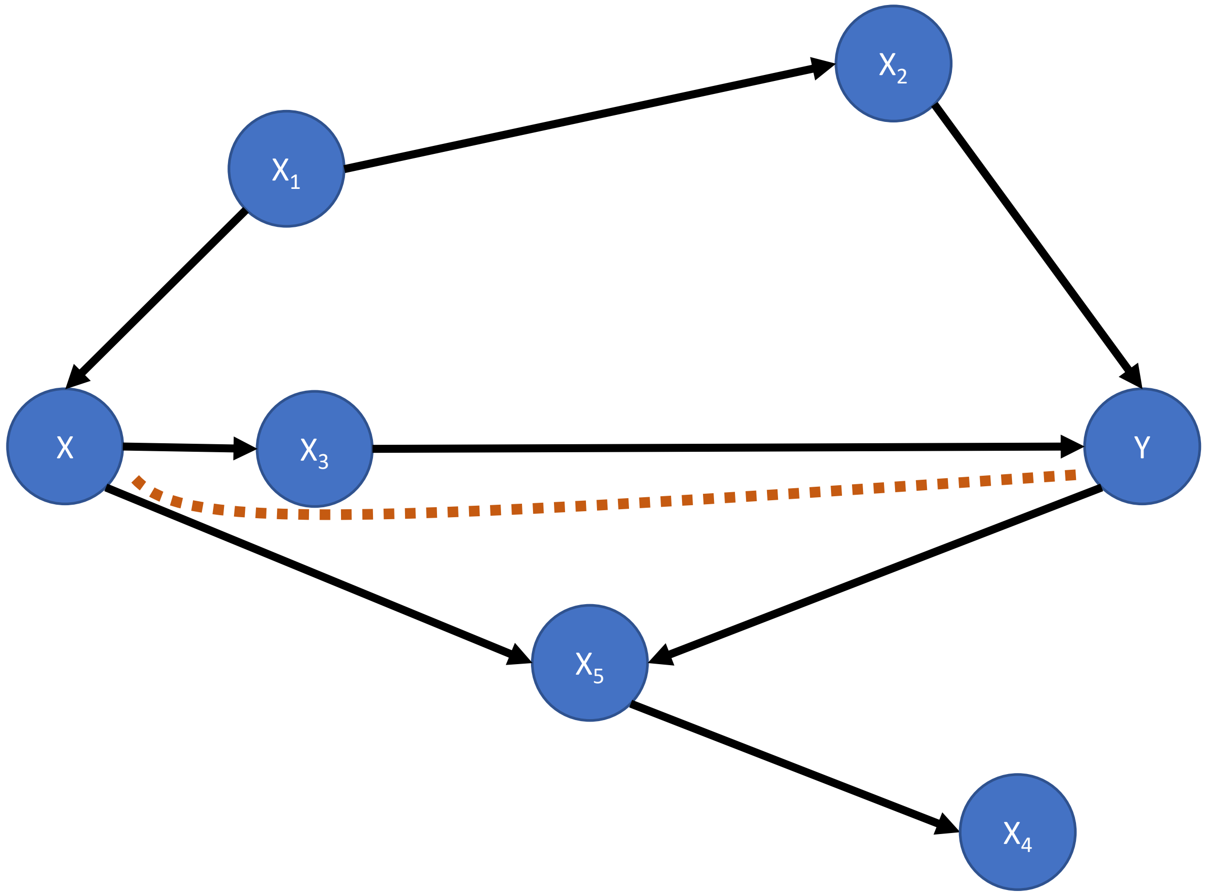
Consider all paths between two nodes as pipes.

- Even if one pipe is unblocked, some water can pass from one node to another.
- To block a pipe, you only need to block it in one place.

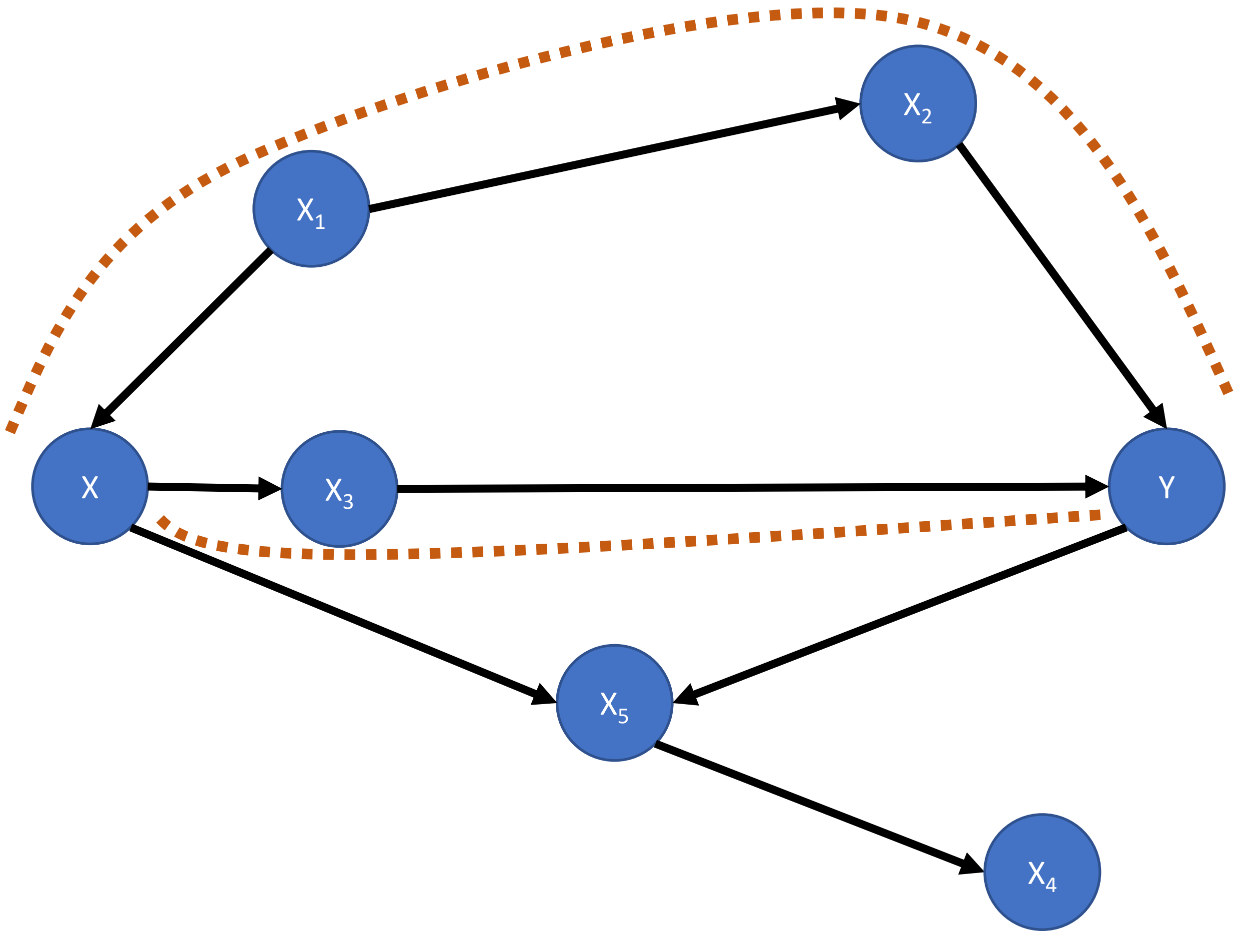
Example



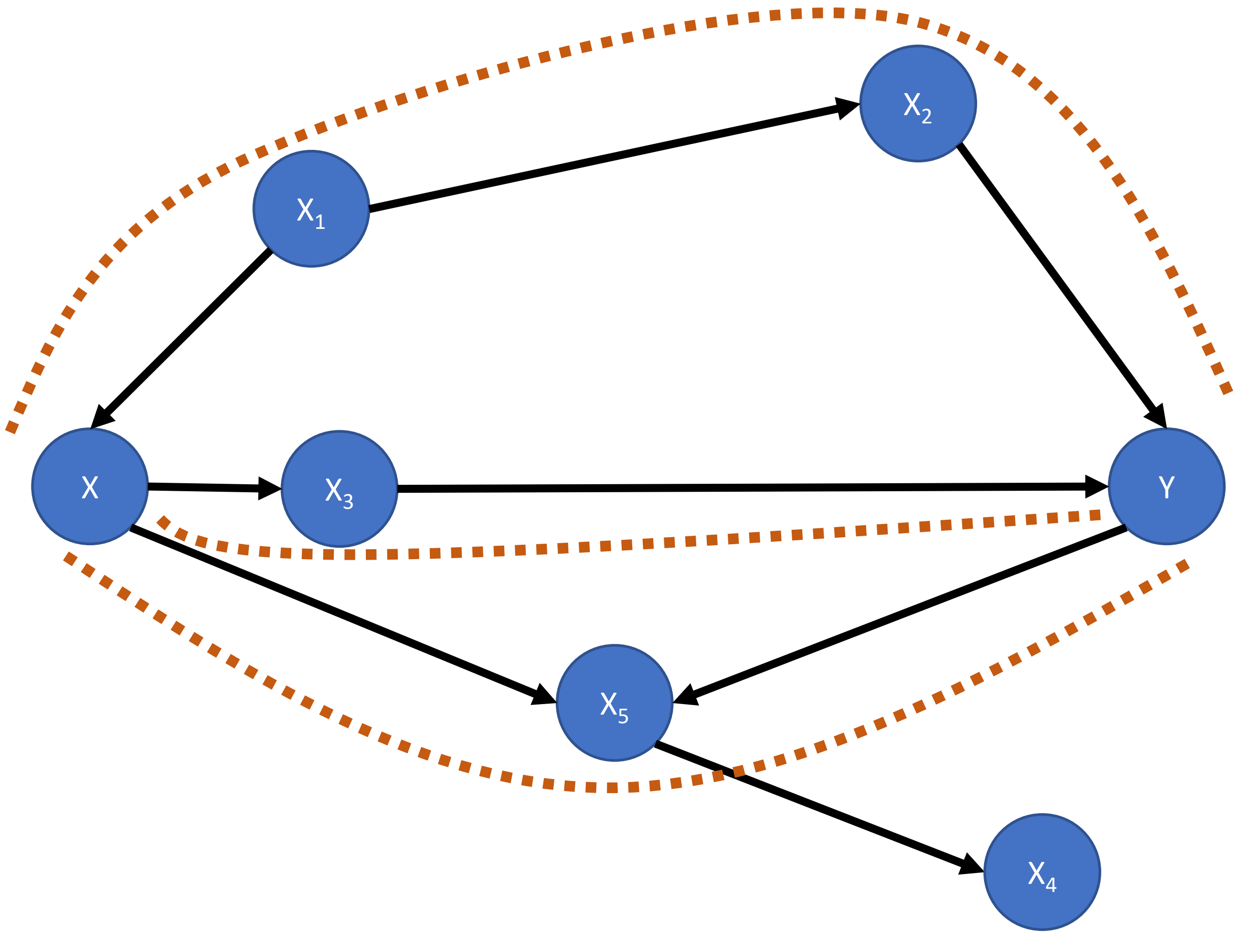
Example



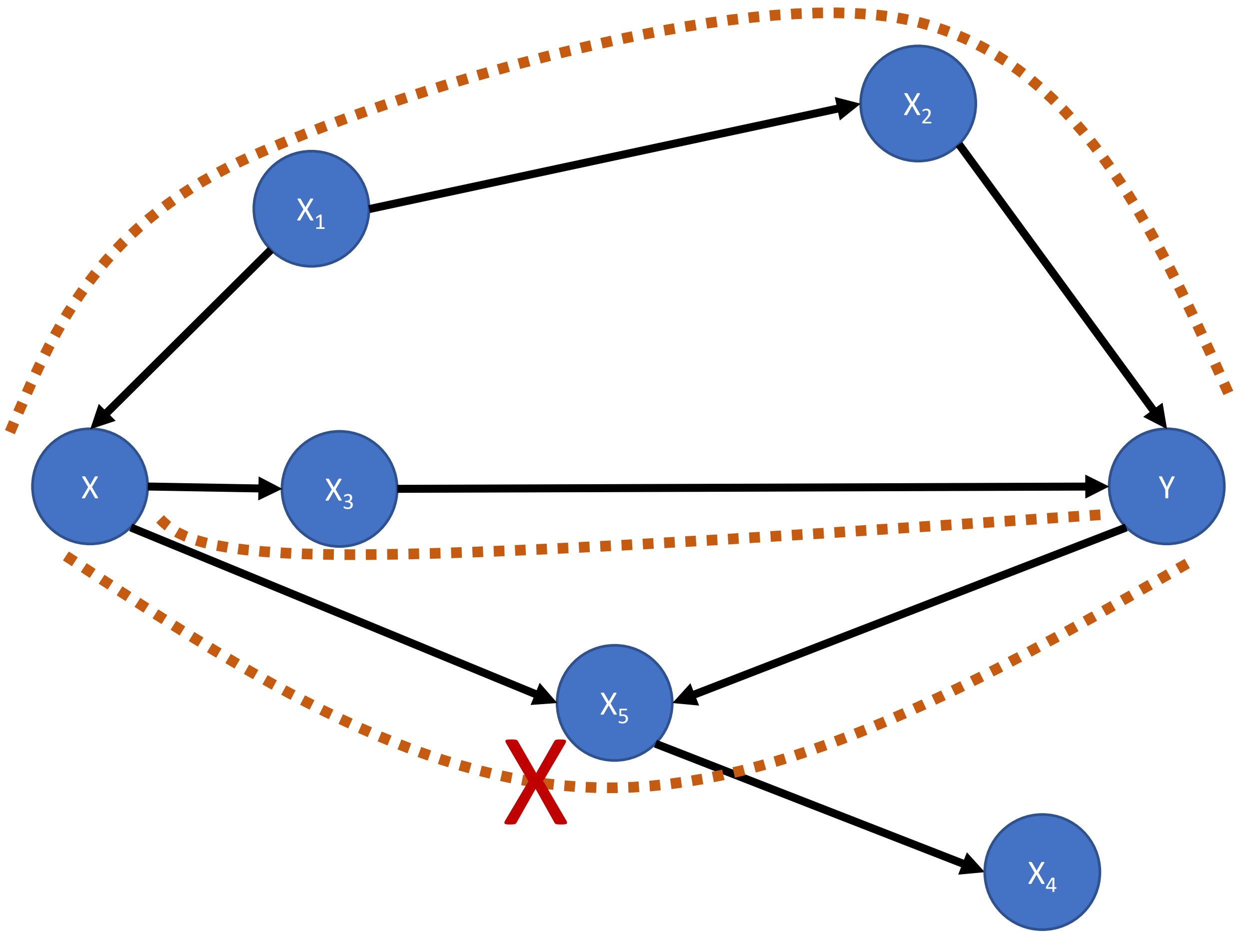
Example



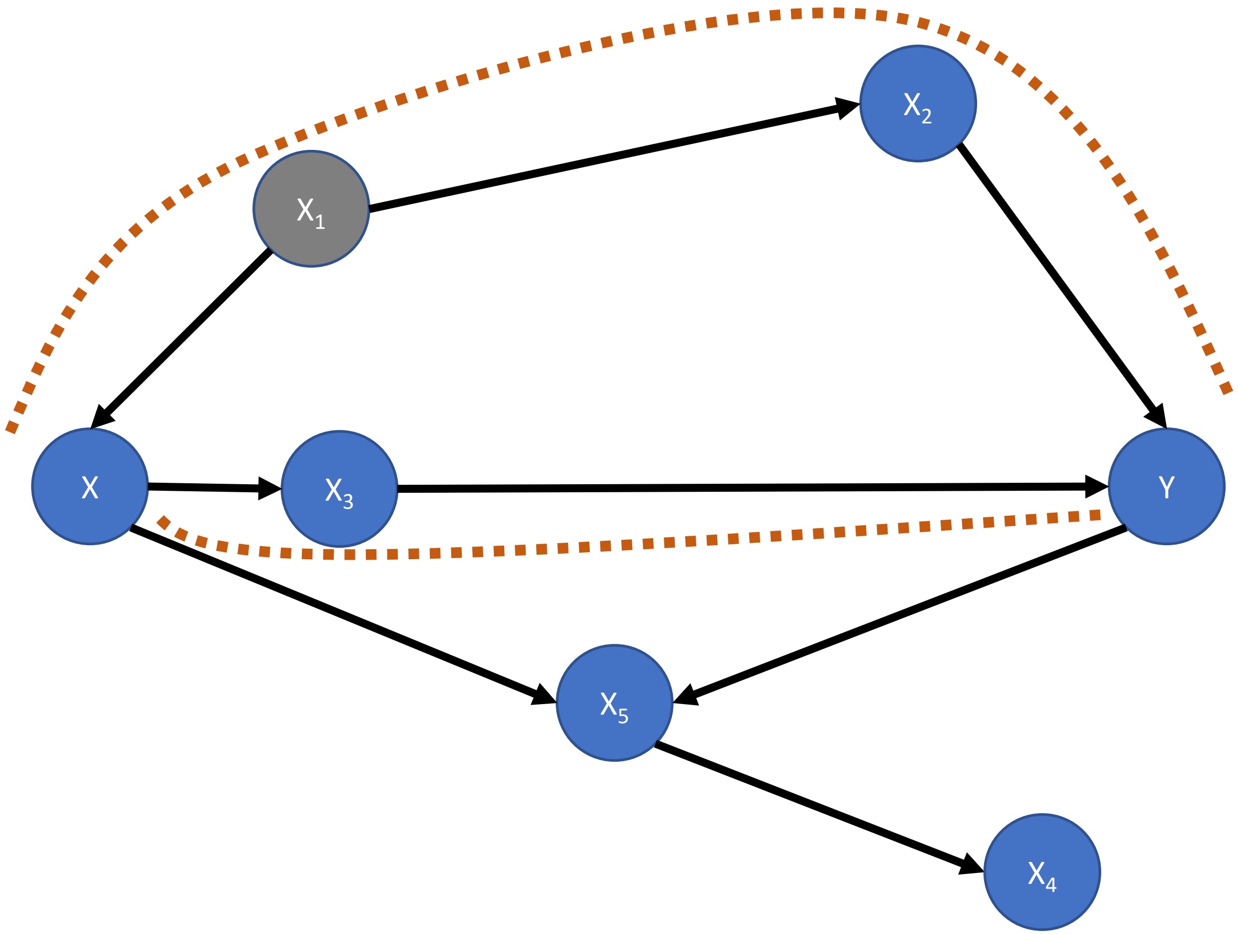
Example



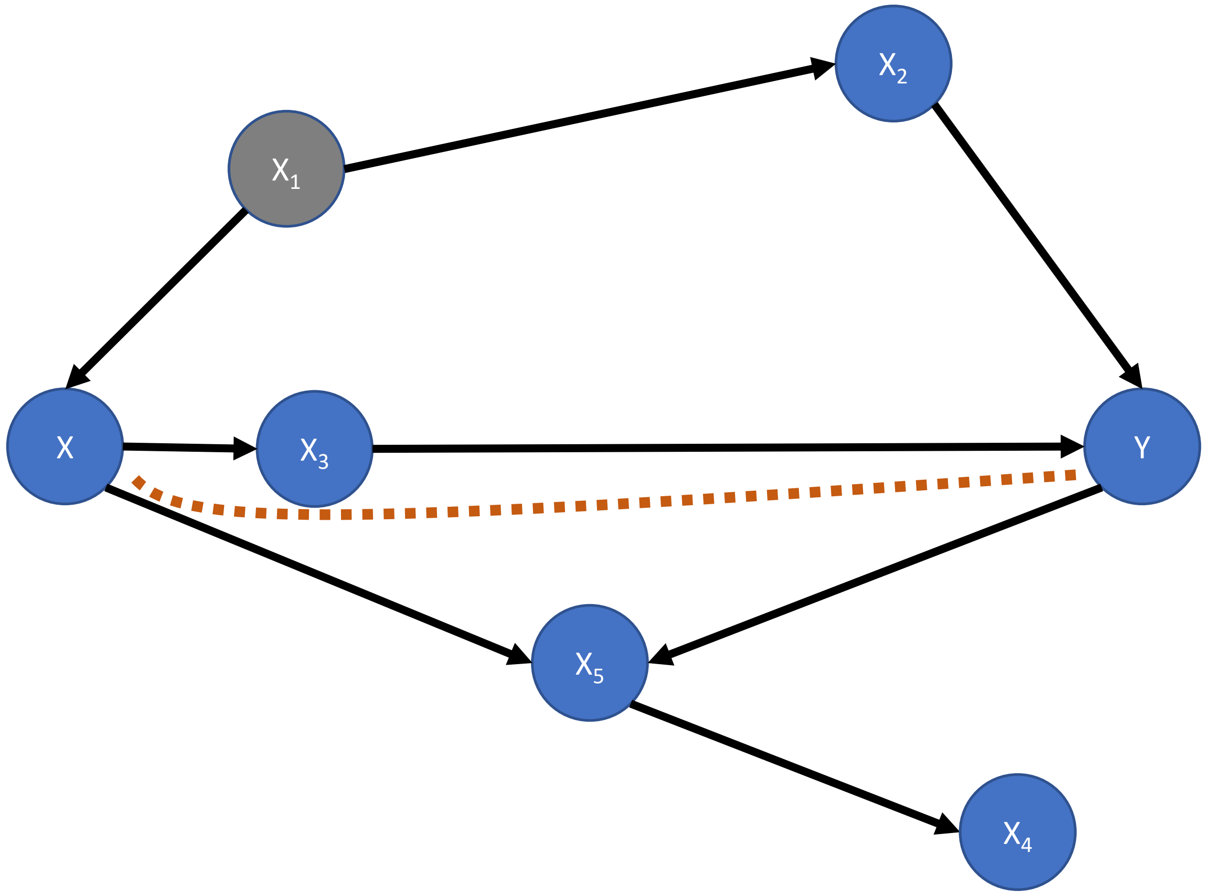
Example



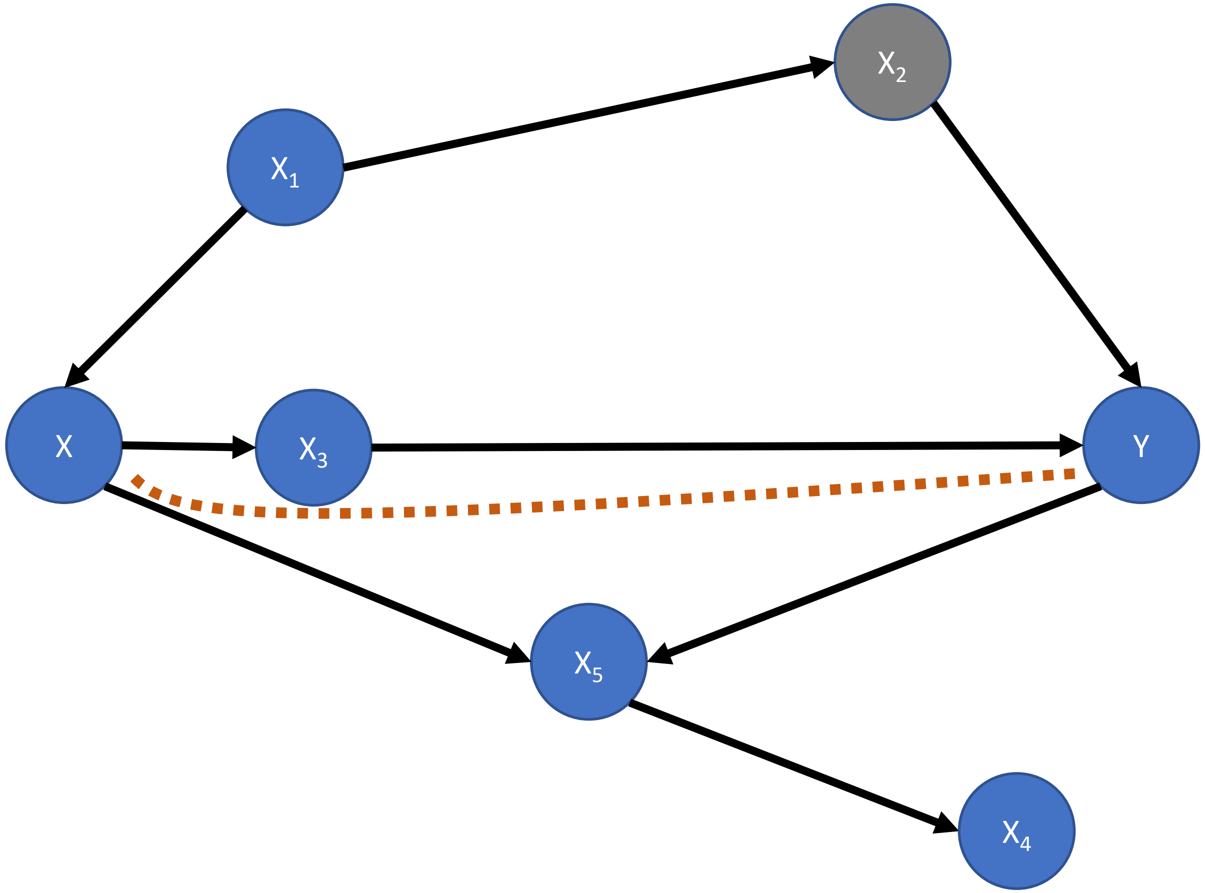
Example



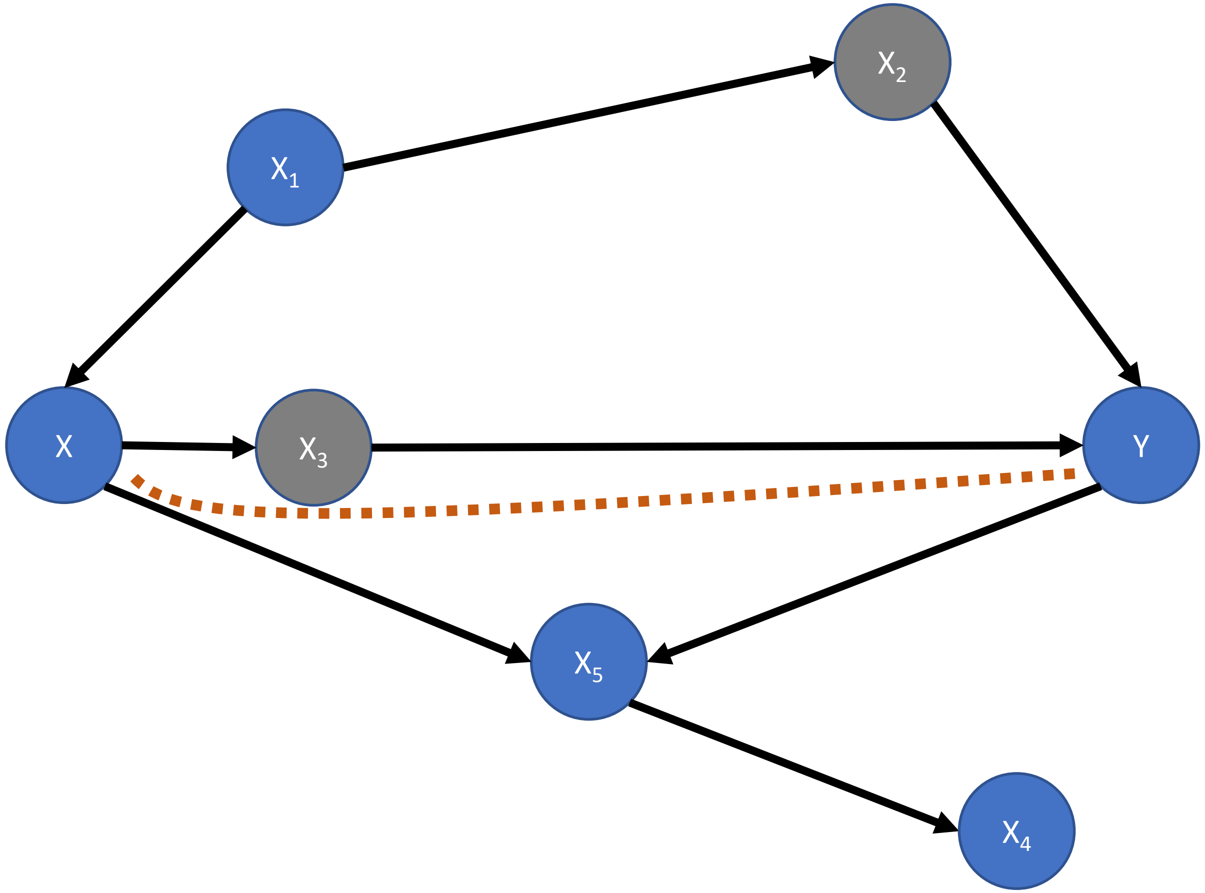
Example



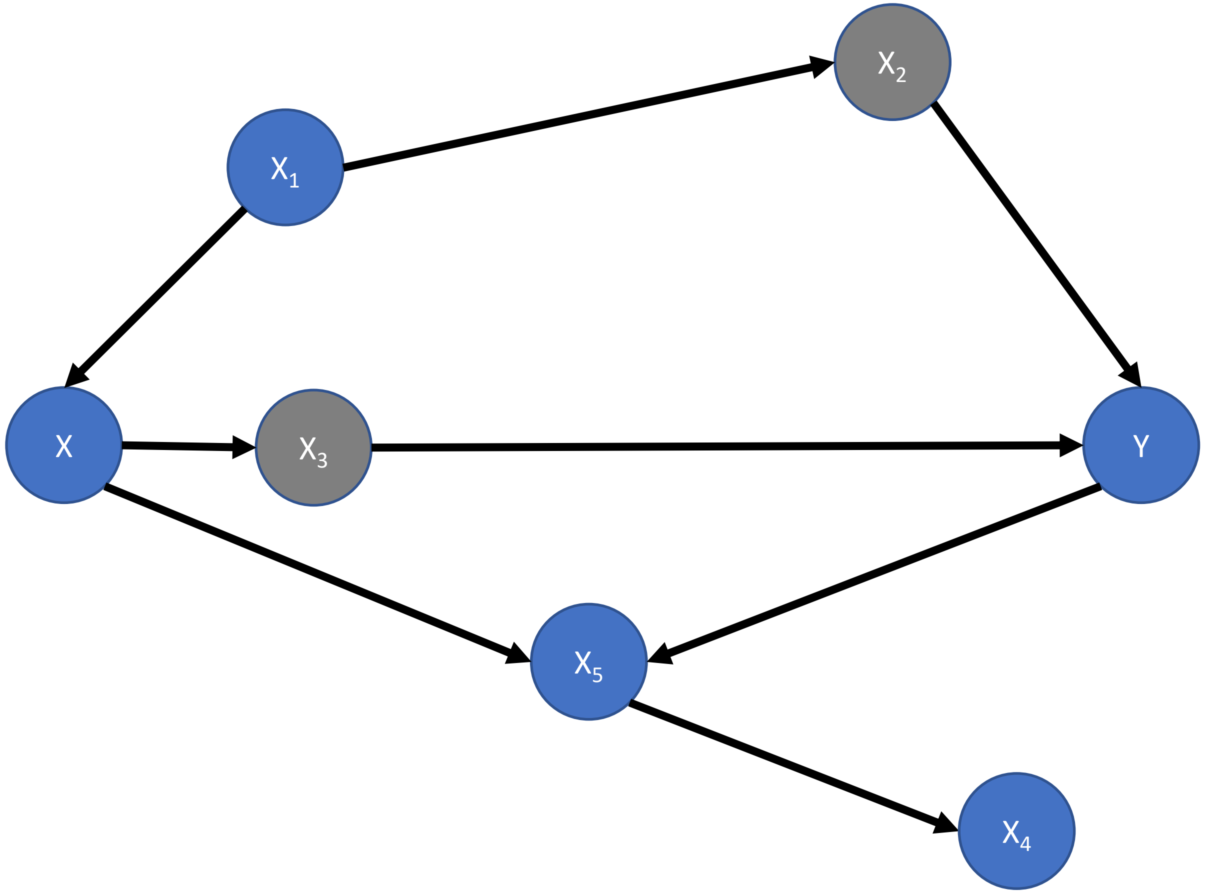
Example



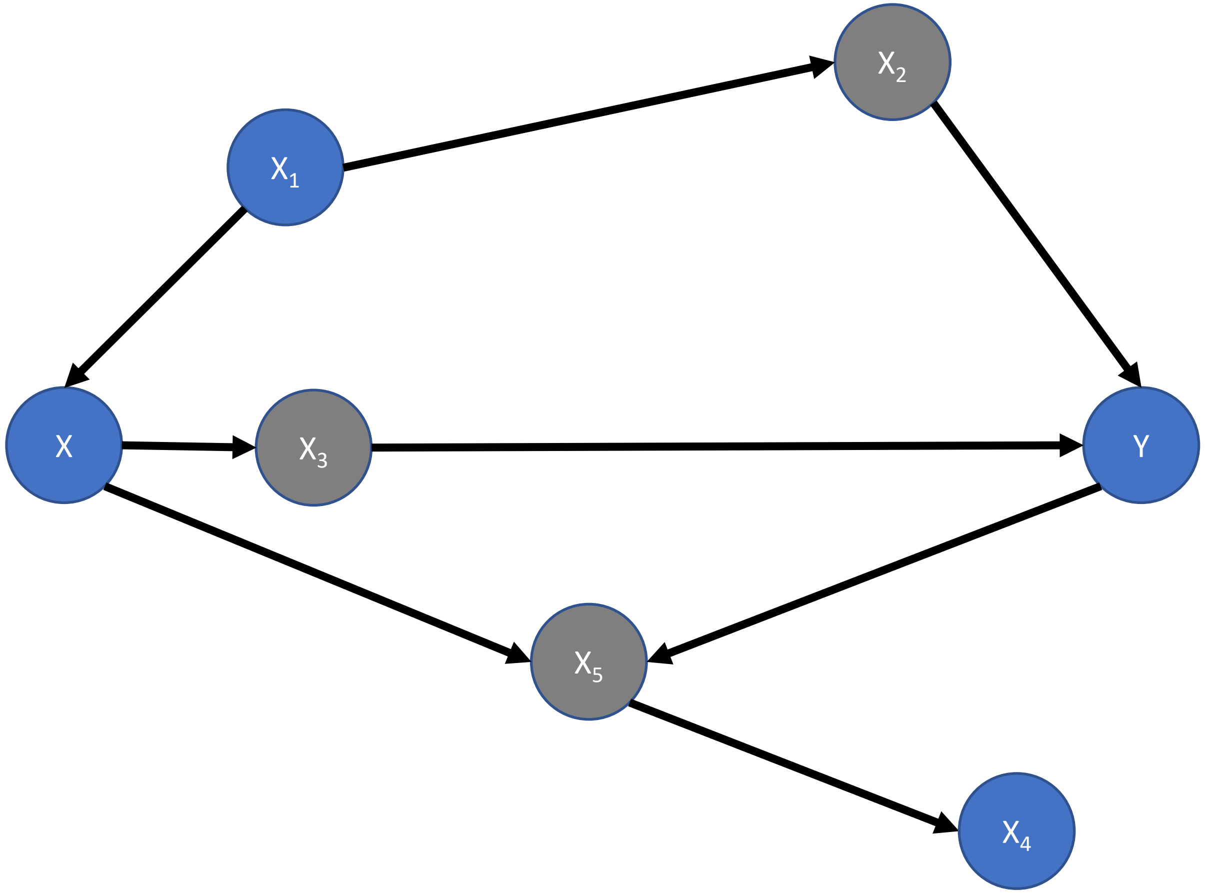
Example



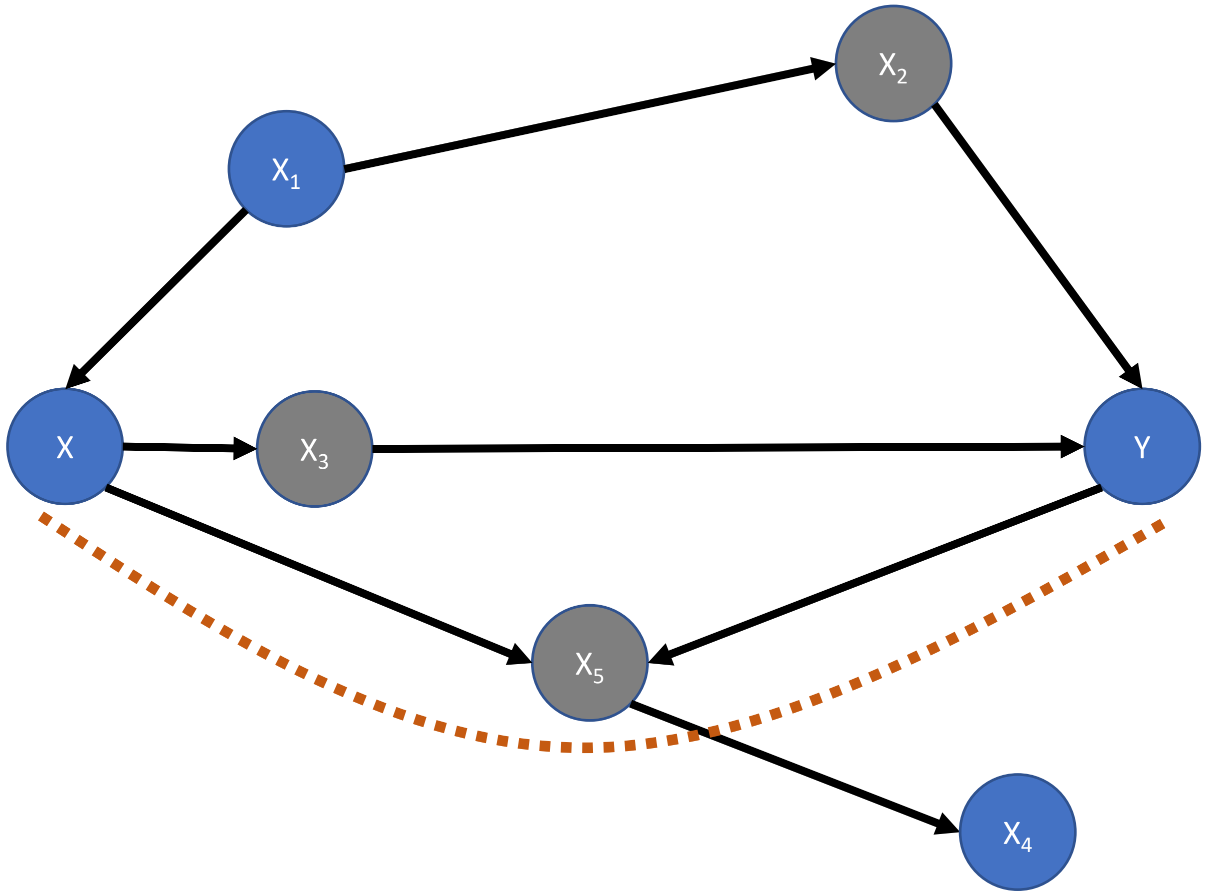
Example



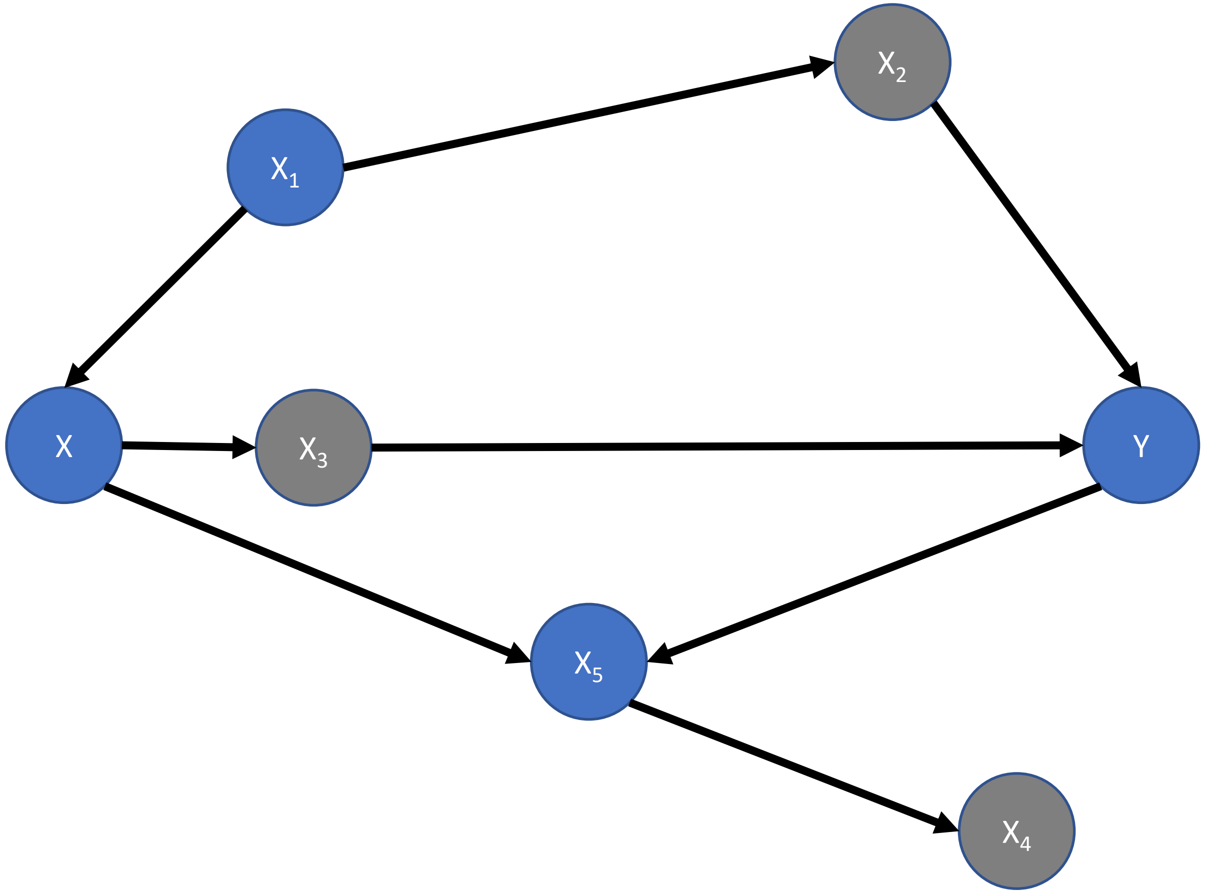
Example



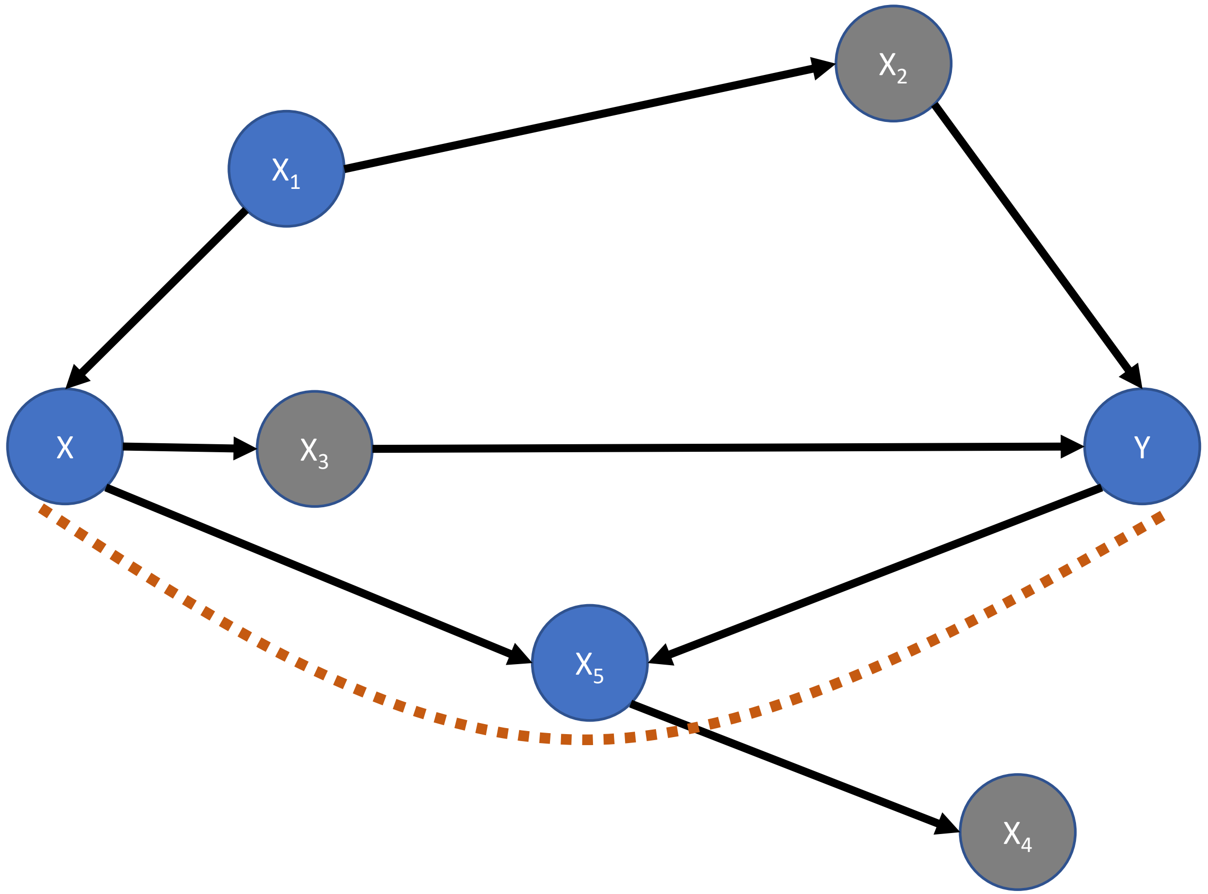
Example



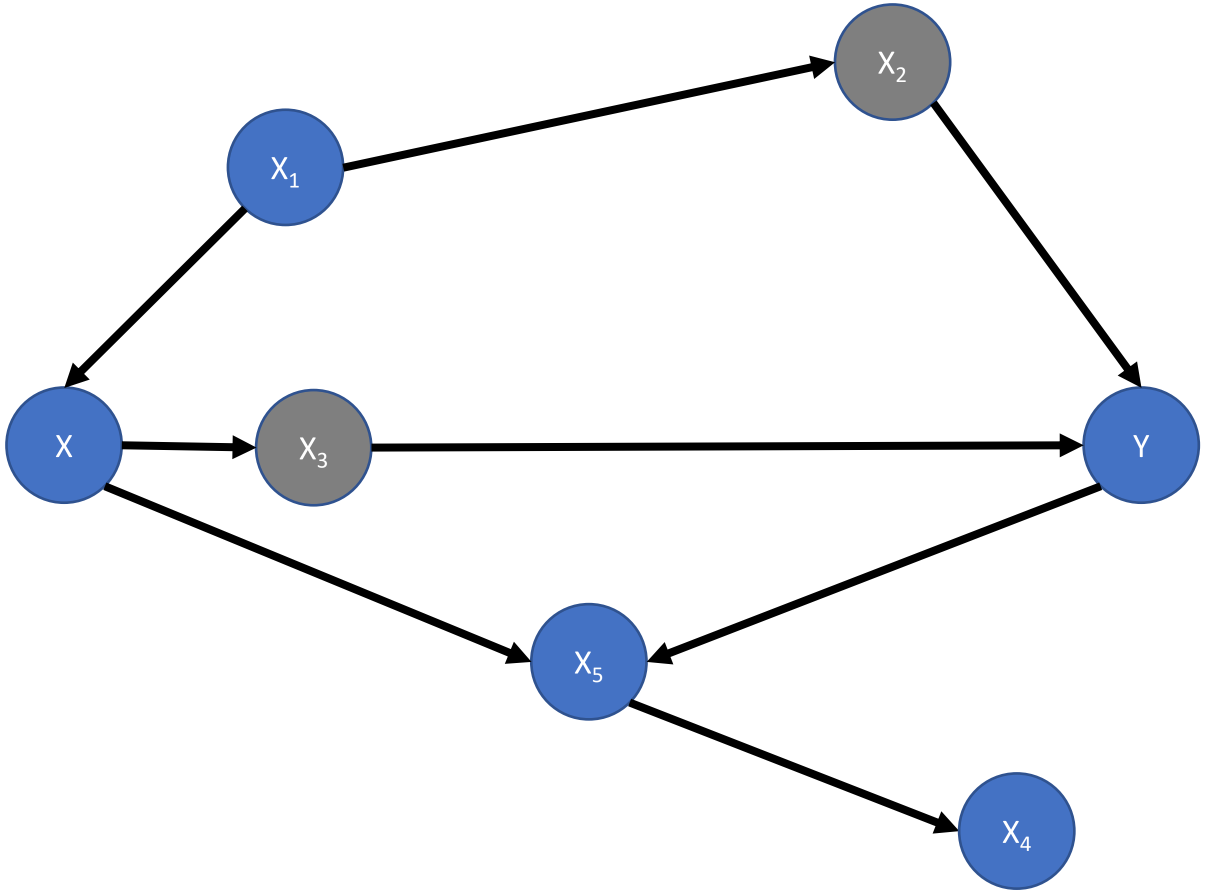
Example



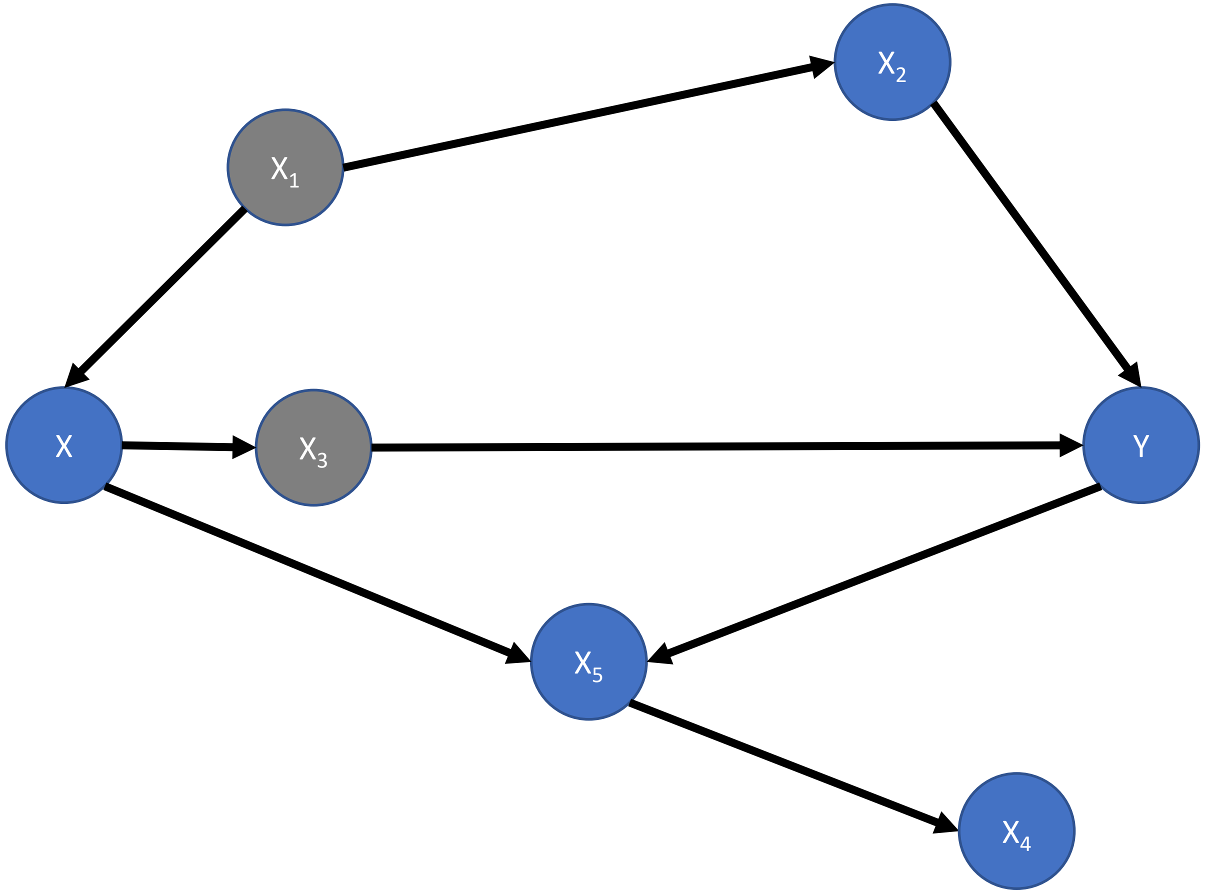
Example



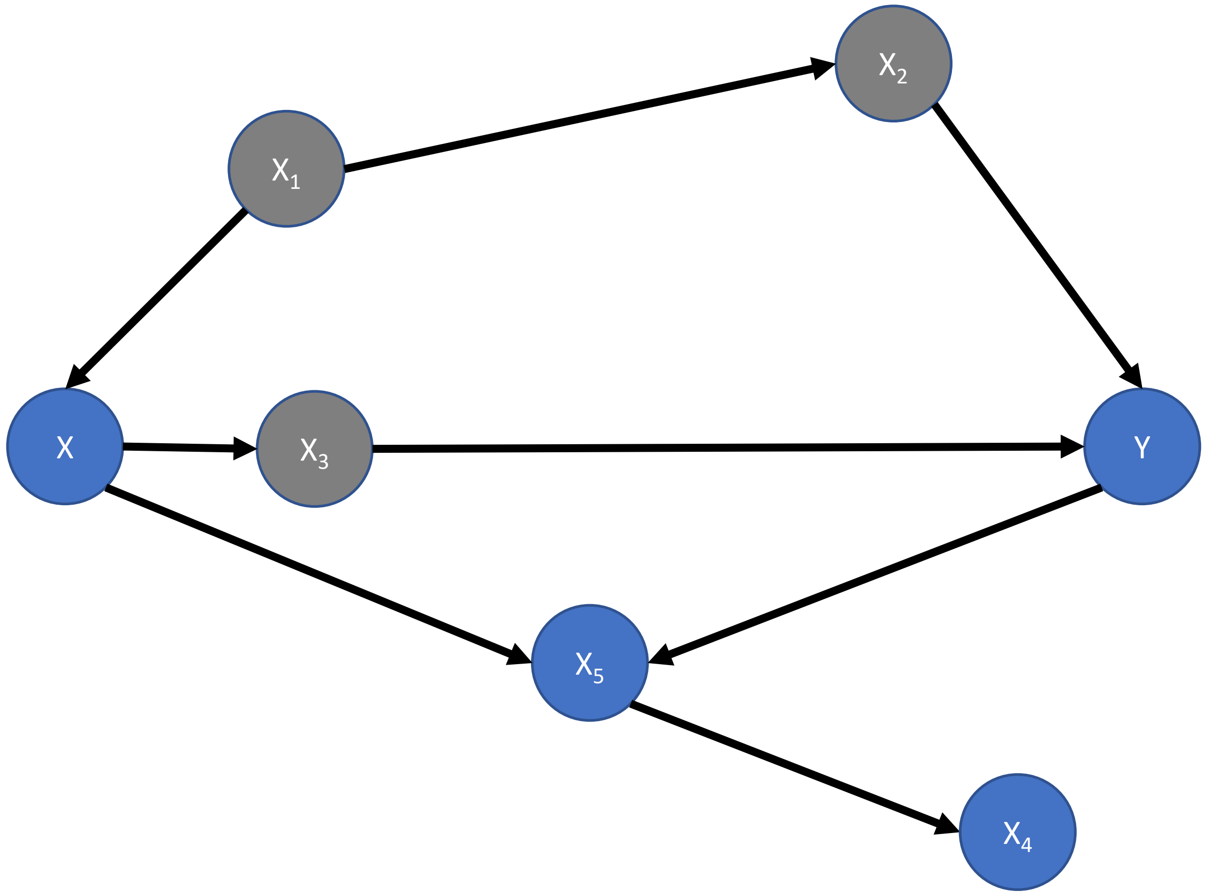
Example: X and Y are d-separated



Example: X and Y are d-separated

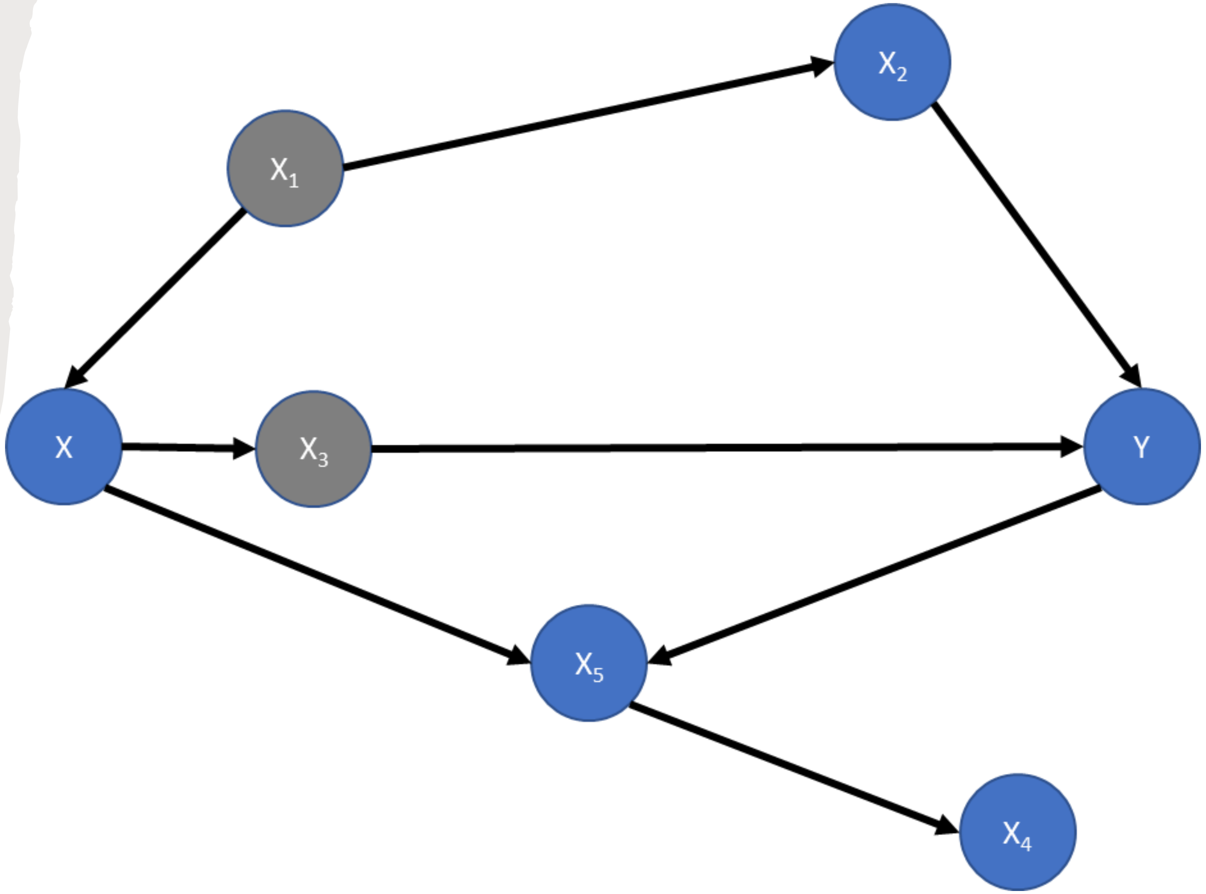


Example: X and Y are d-separated



Model Testing and Causal Discovery

- d-separation can be used to identify statistical implications of the model
  - We can test them!
- $Y=r_X X+ r_{X1} X1+ r_{X3} X3$
  - Y and X are independent, given X1 and X3.
  - $r_X=0$
- If $r_X \neq 0$, model is wrong.
- **<u>Causal Discovery</u>** or Causal Structure Search

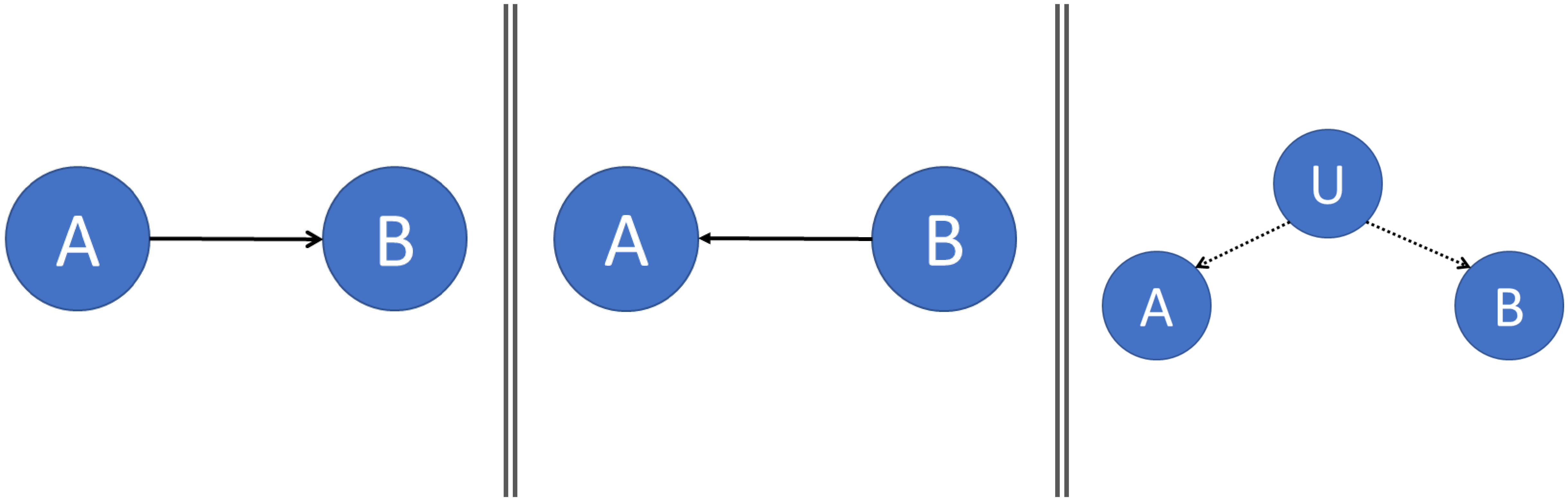


Causal Discovery

Can we learn the DAG from the observed data? No

We need to assume that we have measured all common causes of all variables (Expert Knowledge).

Software tools assume that you have observed all common causes.



Observationally Equivalent but Causally Distinct

Causal effect from Observational data

Assume that the causal model is correct

In some situations, it is possible to “identify” causal effect from observational data.

Association is Causation!

Intervention vs. Conditioning

Intervention: We alter the system

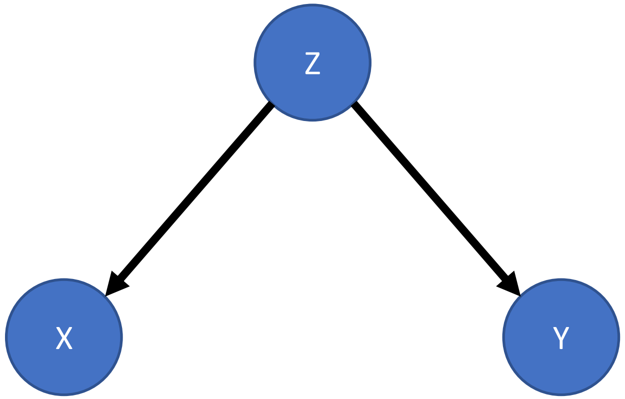
Conditioning: We focus on a subset of data.

Our perception of the system changes not the system

Do-Operator

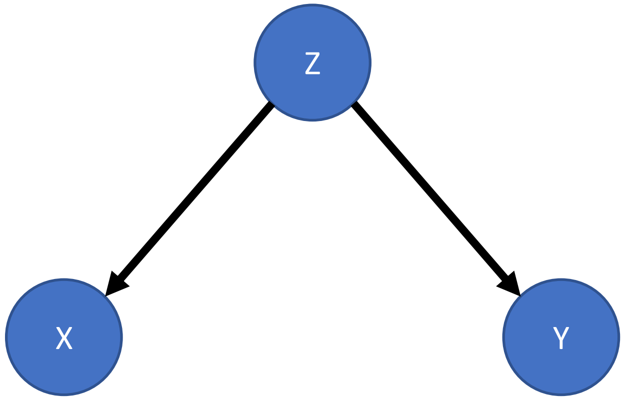
- Intervention: $P(Y = y|do(X = x))$
  - Everyone in the population
  - Causal Effect
- Conditioning: $P(Y = y|X = x)$
  - Subset of population with $X = x$
- $P(Y = y|do(X = x), Z = z)$
  - Both intervention and Conditioning

Do-operation and Graph Manipulation



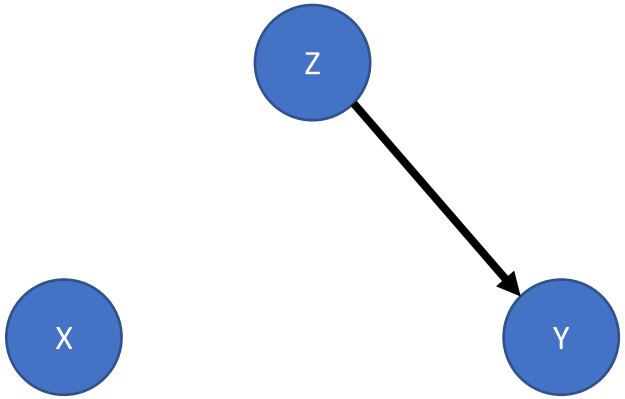
Do-operation and Graph Manipulation

Intervene and do(X=x)



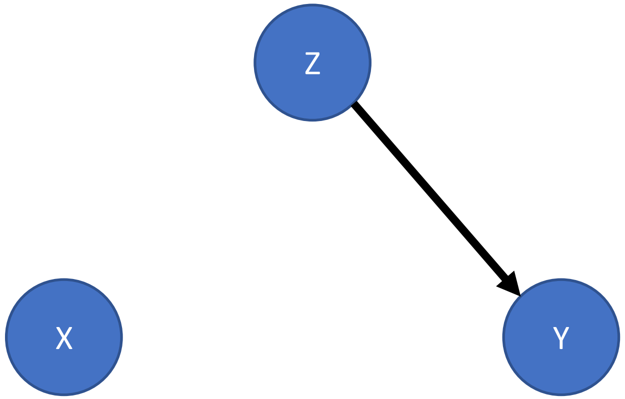
Do-operation and Graph Manipulation

Intervene and do(X=x)



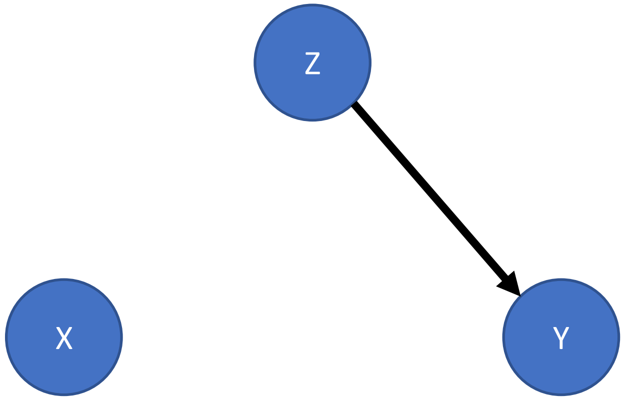
Do-operation and Graph Manipulation

Manipulate the graph and remove all inputs to X



Do-operation and Graph Manipulation

X is not a cause of Y



≡ RCT

Graphical Identification Criteria

In Observational studies, we cannot manipulate the graph

However, we can _sometimes_ **emulate the manipulation.**

Graphical Identification Criteria

- Sets of rules that can be used to check if and how the causal effect is identifiable from the model.

The Backdoor Criterion

- *"Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X"* Pearl, Judea et al. (2016): Causal inference in statistics. A primer.
  - Block all spurious paths: *Backdoors*
  - Leave all directed paths untouched.
  - Don't create any spurious paths

$$P(Y = y | do(X = x)) = \sum_z P(Y = y | X = x, Z = z) P(Z = z)$$

The Backdoor Criterion

- *"Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X"* Pearl, Judea et al. (2016): Causal inference in statistics. A primer.
  - Block all spurious paths: *Backdoors*
  - Leave all directed paths untouched.
  - Don't create any spurious paths

Adjustment Formula

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

Inverse Probability Weighting

$$P(Y = y|do(X = x)) = \sum_z \frac{P(Y=y, X=x, Z=z)}{P(X=x|Z=z)}$$

Inverse Probability Weighting

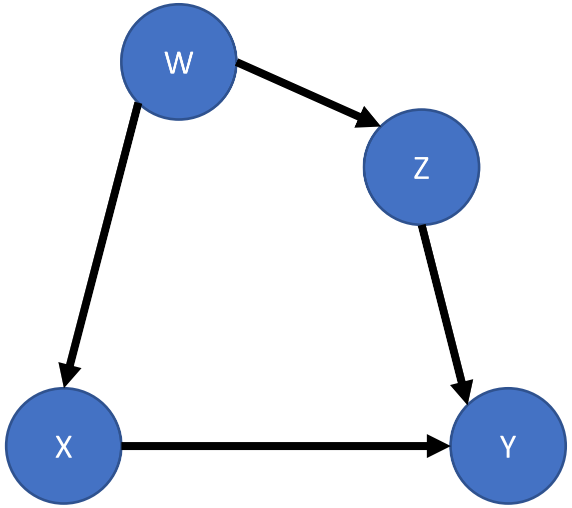
$$P(Y = y|do(X = x)) = \sum_z \frac{P(Y=y, X=x, Z=z)}{P(X=x|Z=z)}$$

Inverse Probability Weighting

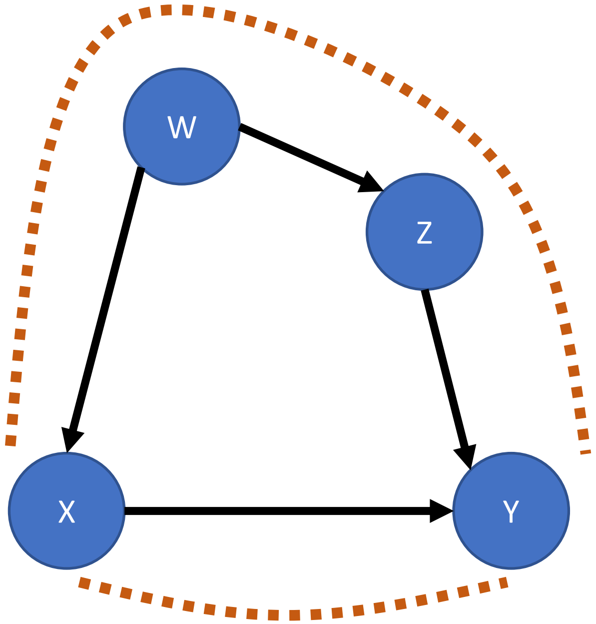
$$P(Y = y|do(X = x)) = \sum_z \frac{P(Y=y,X=x,Z=z)}{P(X=x|Z=z)}$$

Propensity Score

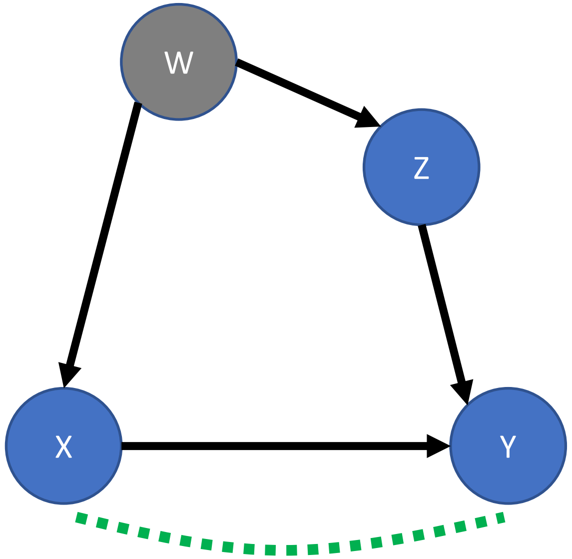
Example: Causal Effect of X on Y



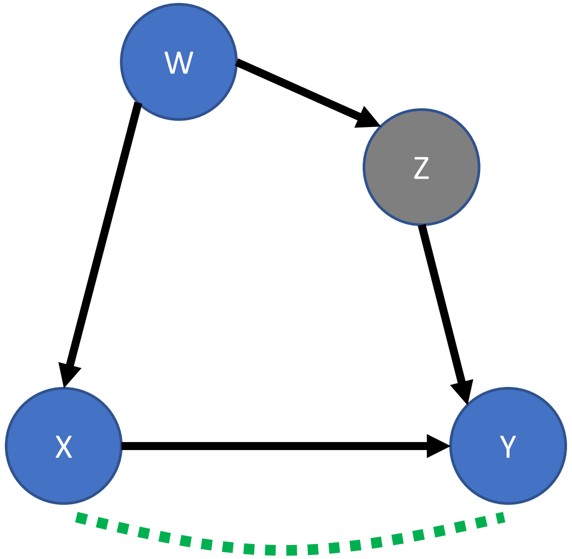
Example



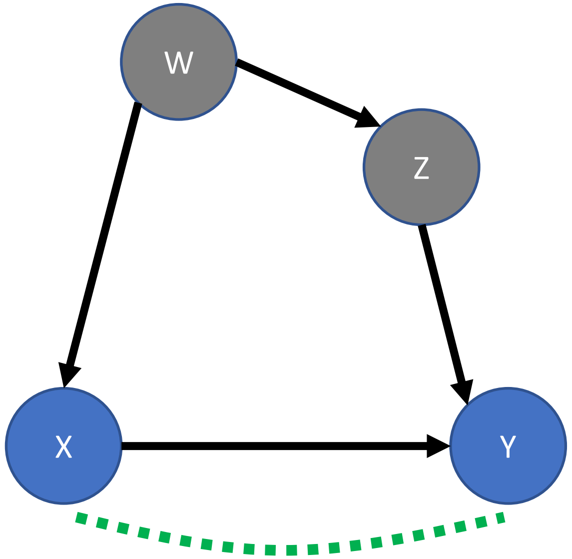
Example



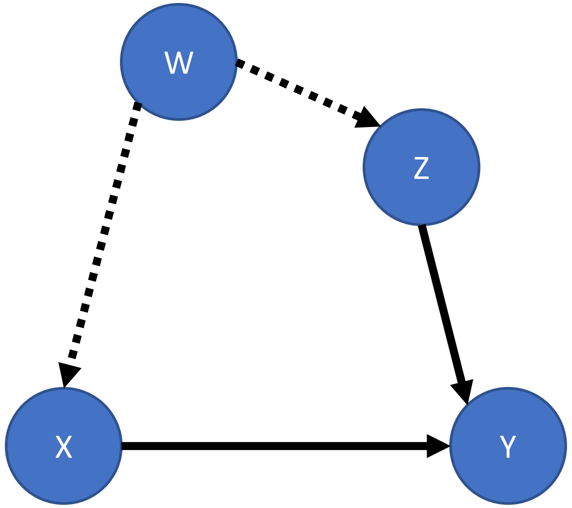
Example



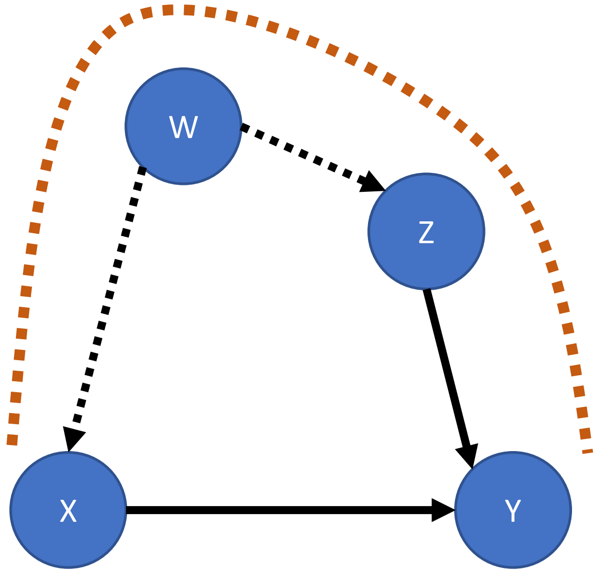
Example



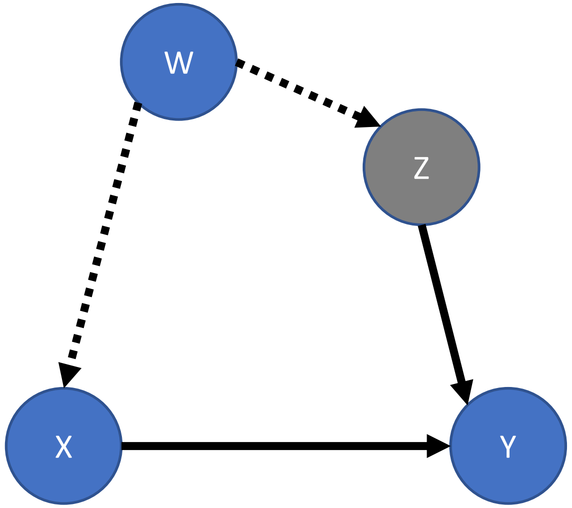
Example



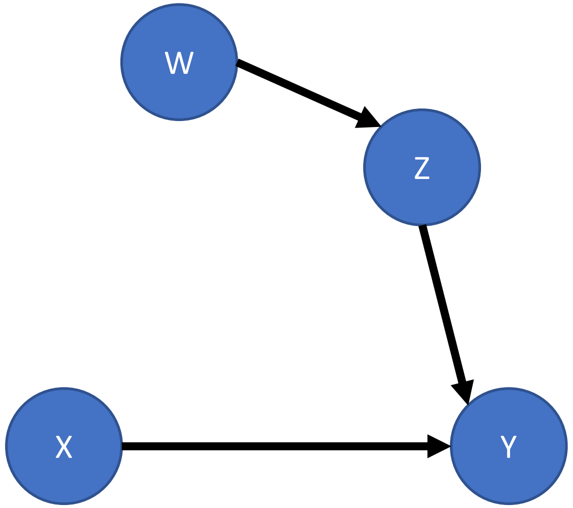
Example



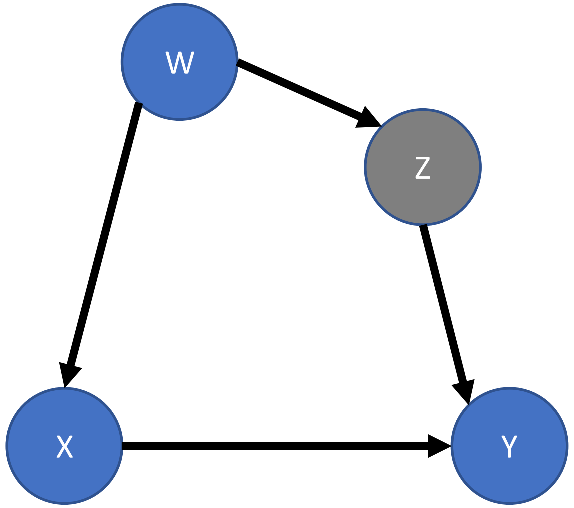
Example



Example

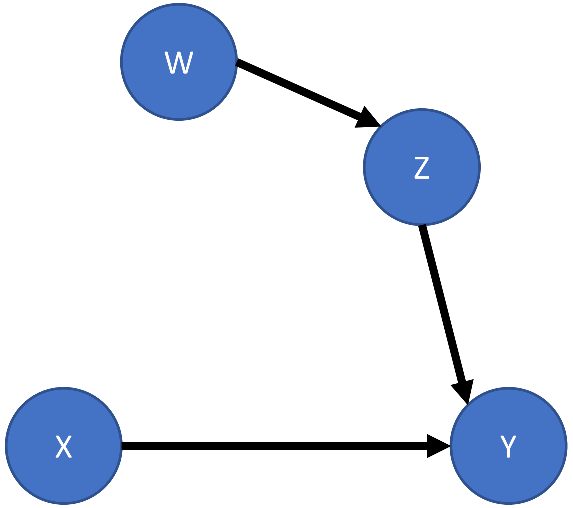


$$P(Y = y | do(X = x))$$

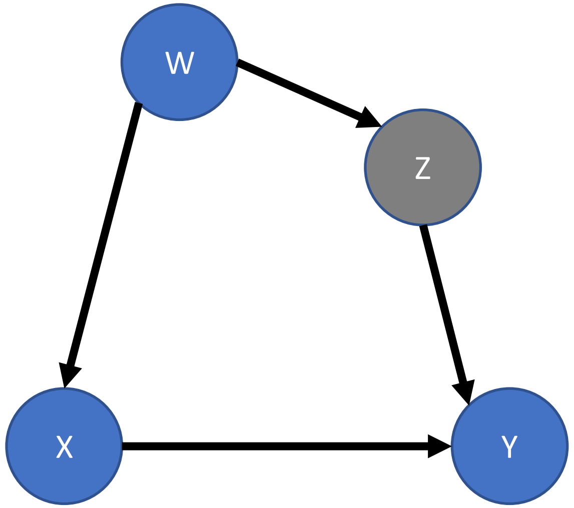


$$\sum_z P(Y = y | X = x, Z = z) P(Z = z)$$

Example

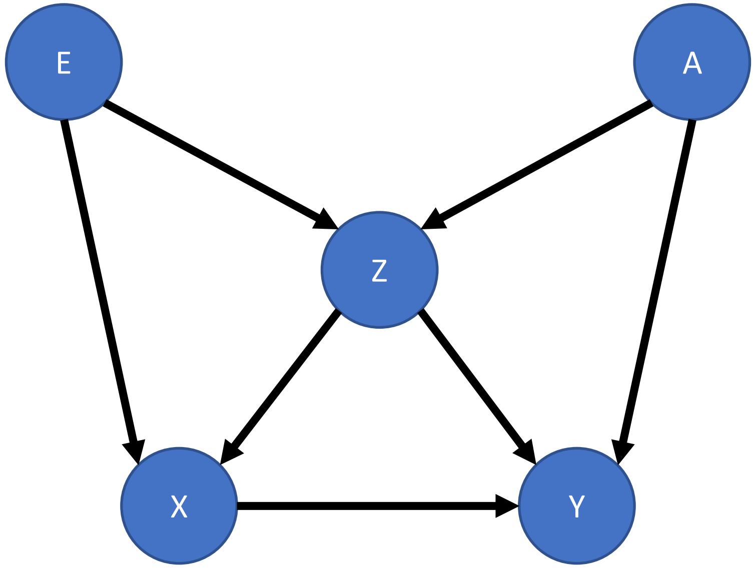


$P(Y = y|do(X = x))$



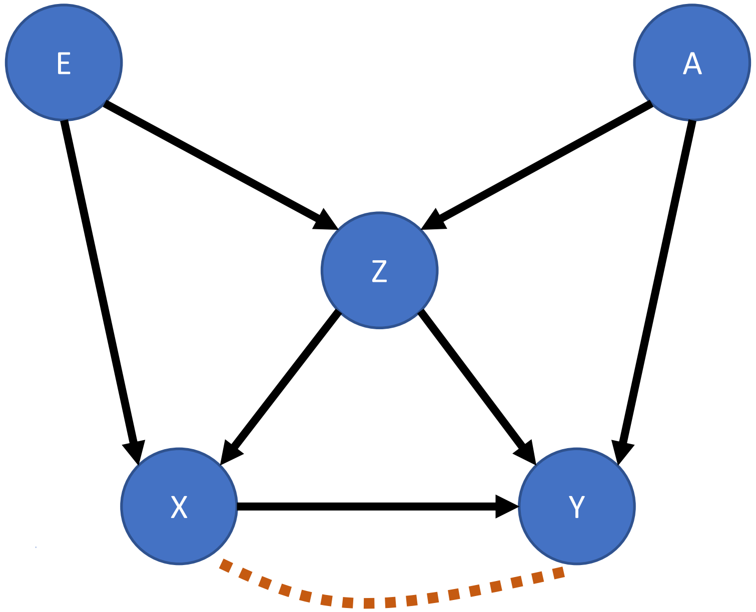
$\sum_z P(Y = y|X = x, Z = z)P(Z = z)$

Example

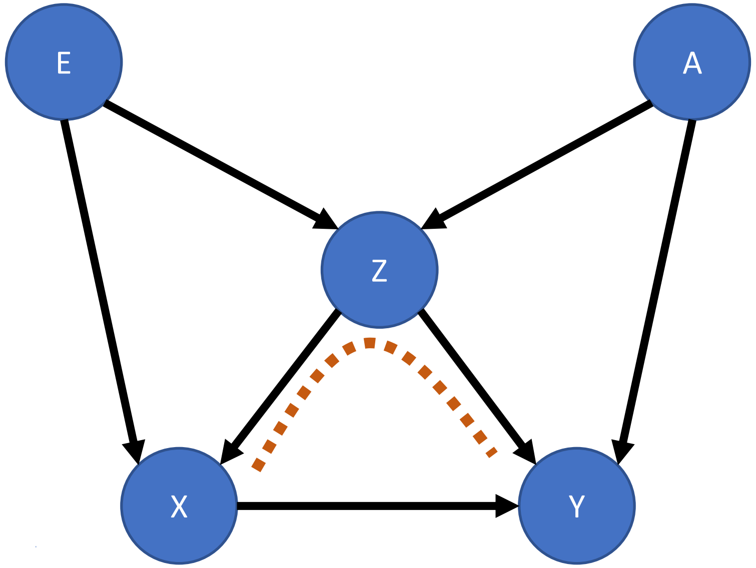


Pearl, Judea et al. (2016): Causal inference in statistics. A primer.

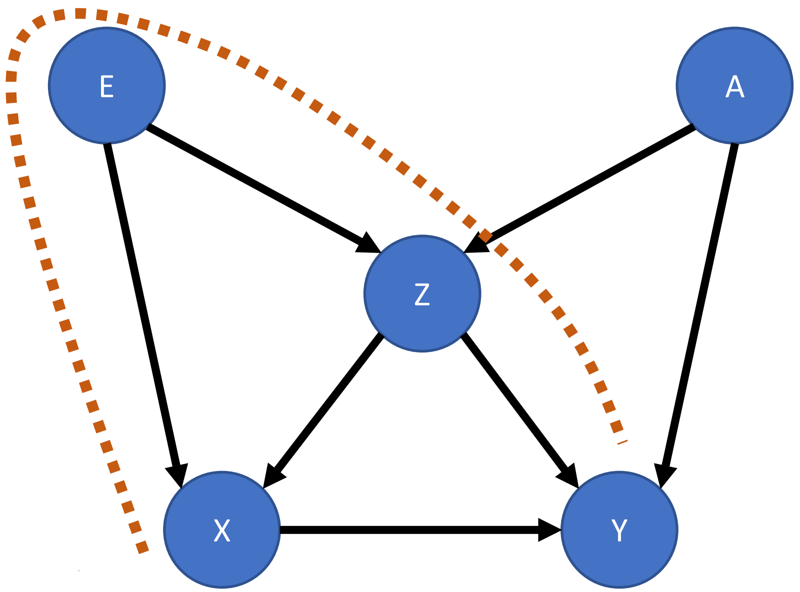
Example



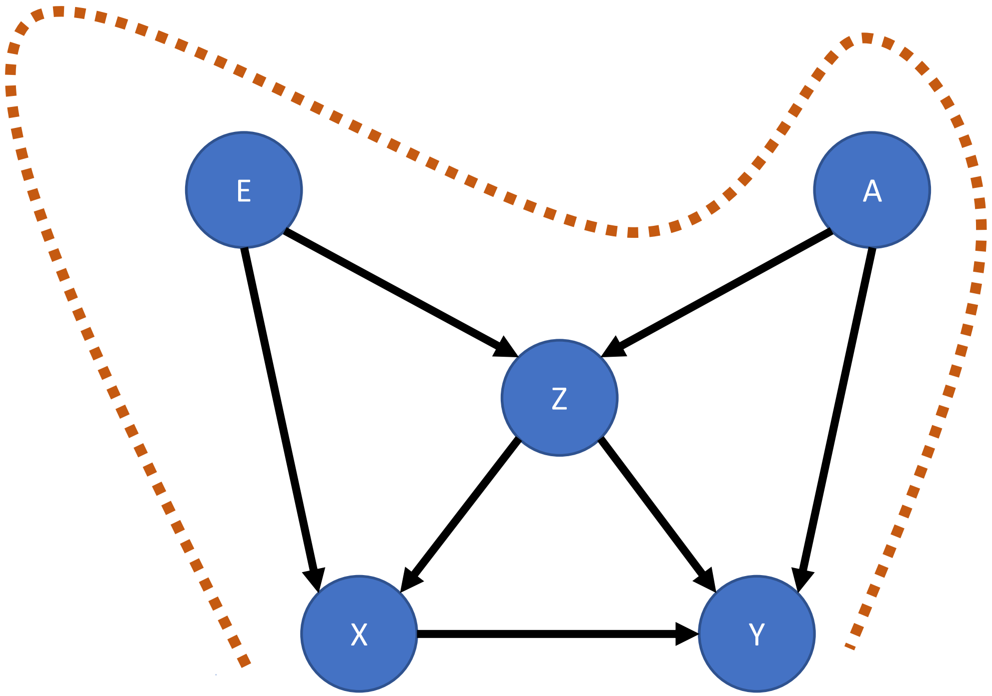
Example



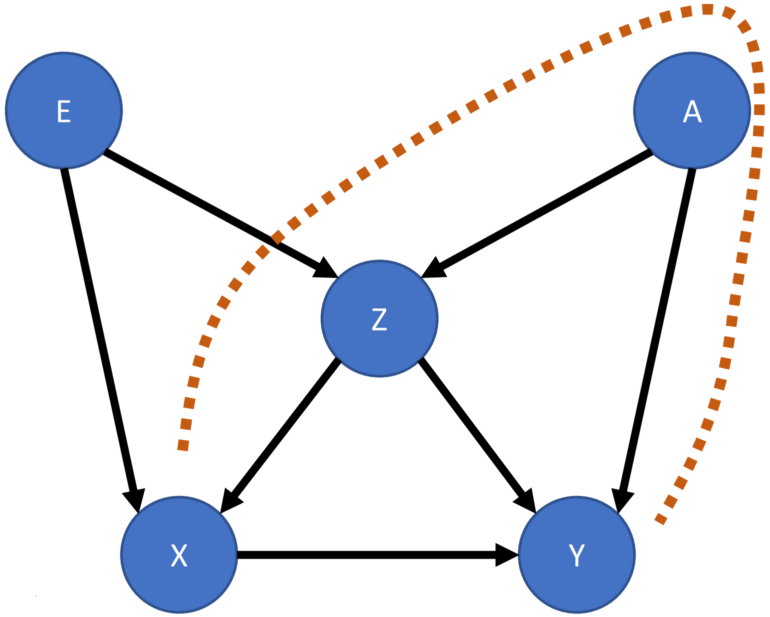
Example



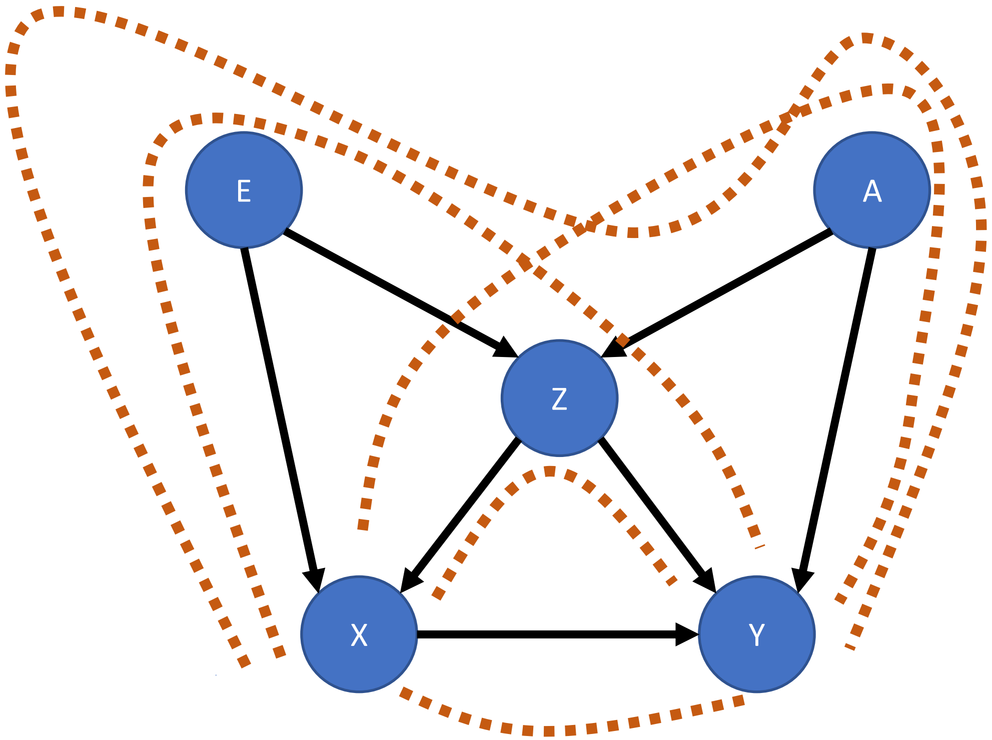
Example



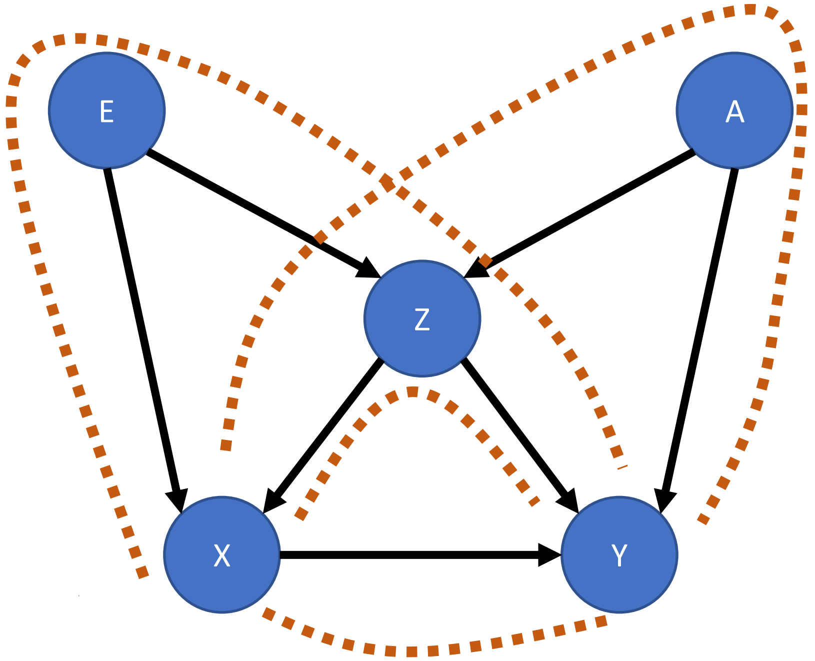
Example



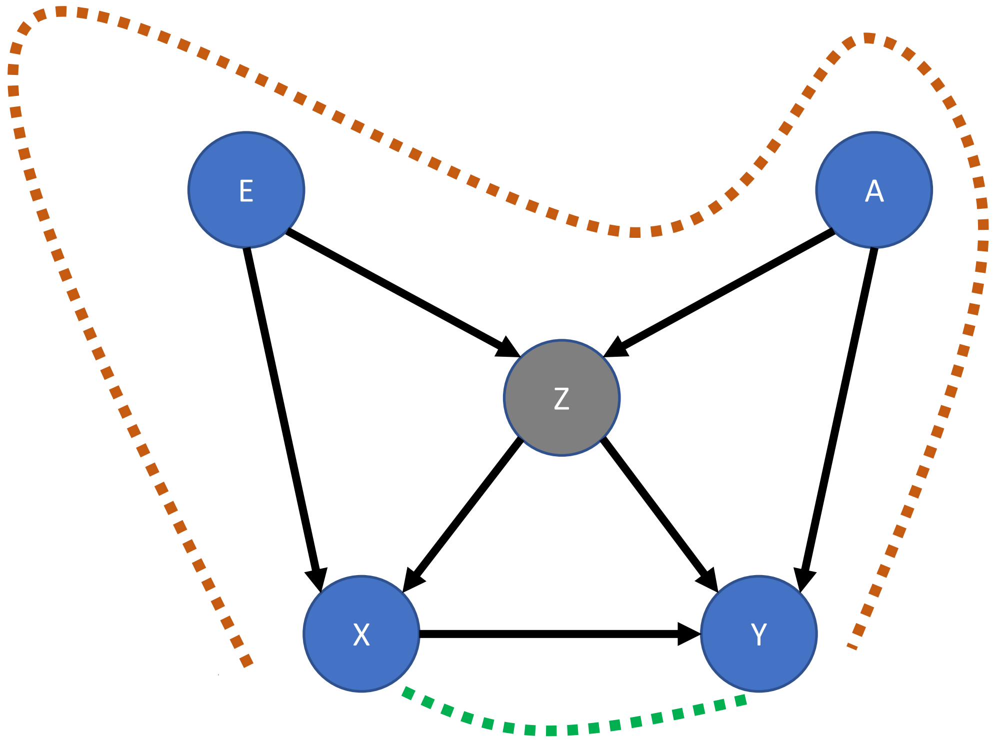
Example



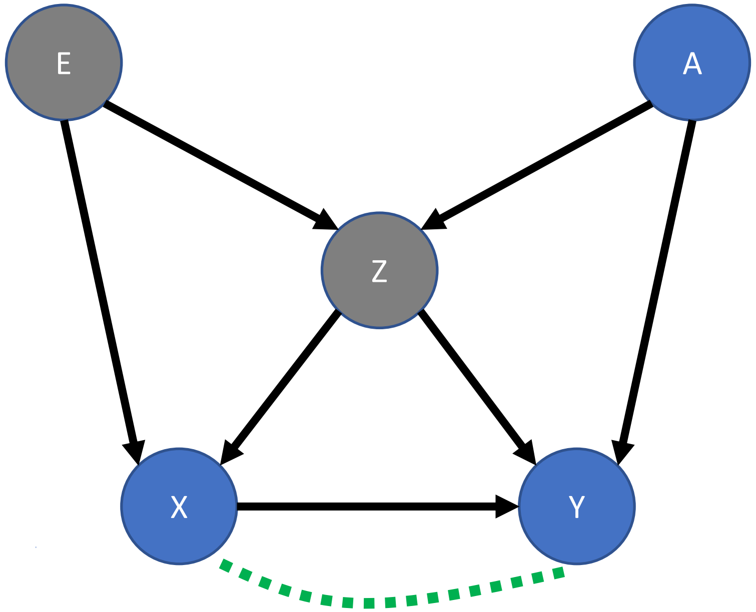
Example



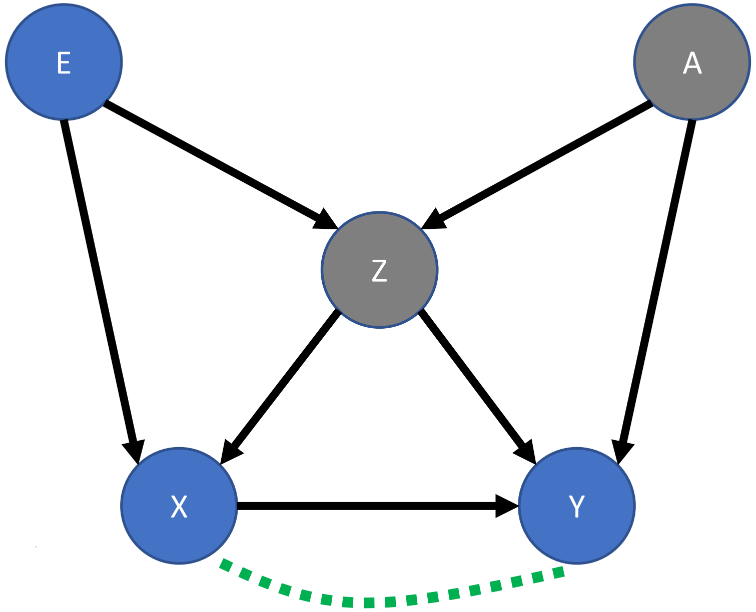
Example



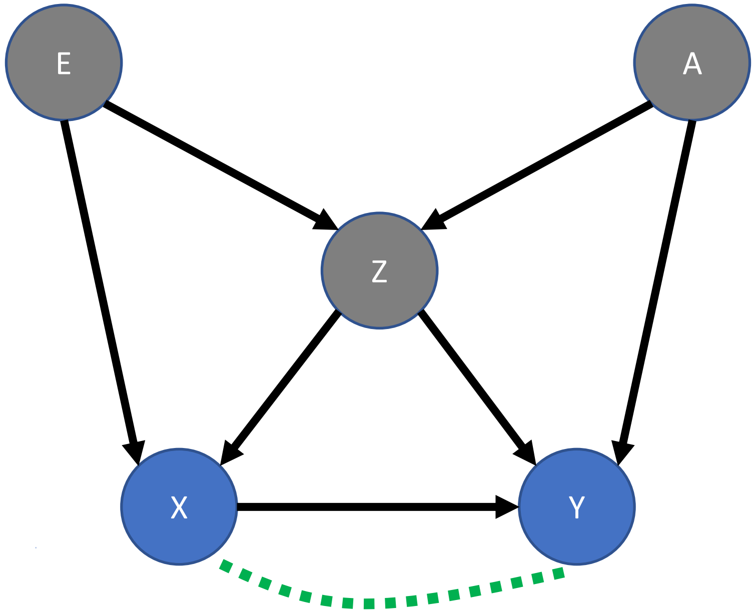
Example



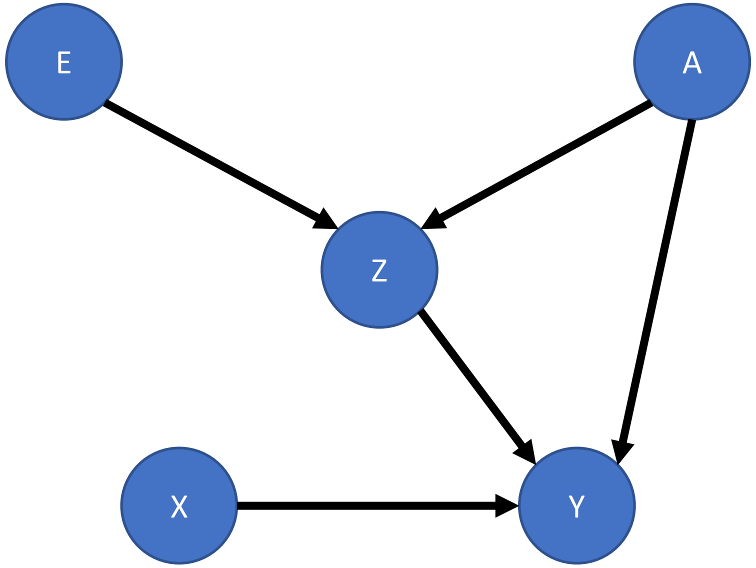
Example



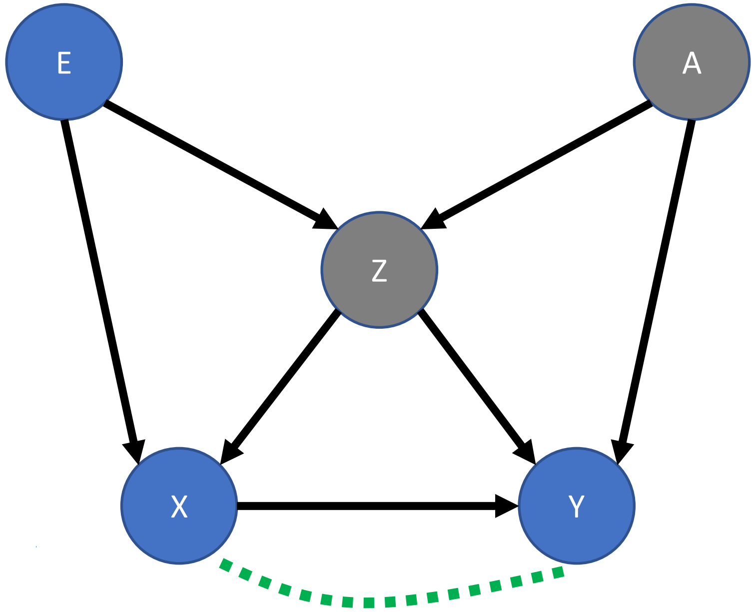
Example



Example



$$P(Y = y | do(X = x))$$



$$\sum_z P(Y = y | X = x, Z = z, A = a) P(Z = z, A = a)$$

Do-calculus and identifiability of Causal Estimand

Backdoor criterion is a sufficient criterion.

There are other criteria that can be used such as Front-door Criterion

- It is also a sufficient criterion.

Do-Calculus rules solve this problem. If there is a way to identify a causal effect, we can find it.

- Necessary and Sufficient

Some Examples From Literature

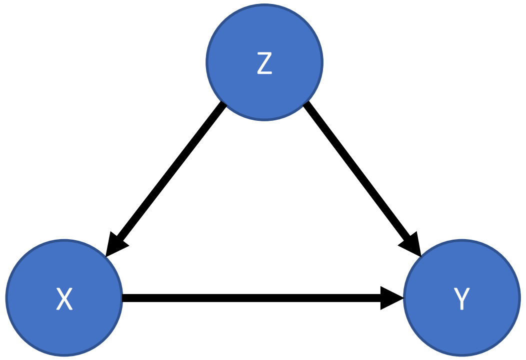
Simpson's Paradox

Treatment	Male	Female	Total
Yes	81/87 **(93%)**	192/263 **(73%)**	273/350 (78%)
No	234/270 (87%)	55/80 (69%)	289/350 **(83%)**

Pearl, Judea et al. (2016): Causal inference in statistics. A primer.

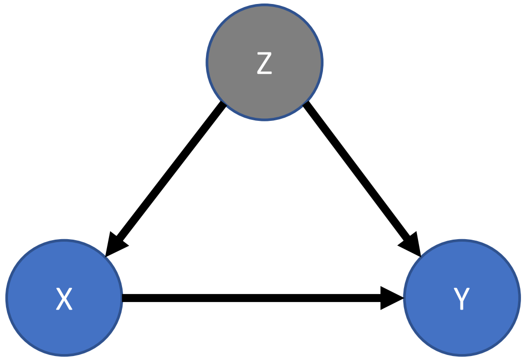
Simpson's Paradox

Treatment	Male	Female	Total
Yes	81/87 **(93%)**	192/263 **(73%)**	273/350 (78%)
No	234/270 (87%)	55/80 (69%)	289/350 **(83%)**



Simpson's Paradox

Treatment	Male	Female	Total
Yes	81/87 **(93%)**	192/263 **(73%)**	273/350 (78%)
No	234/270 (87%)	55/80 (69%)	289/350 **(83%)**



$$P(Y = yes|do(X = yes)) = \sum_{z=\{male,female\}} \frac{P(Y = yes, X = yes, Z = z)}{P(X = yes|Z = z)}$$

$$P(Y = yes|do(X = no)) = \sum_{z=\{male,female\}} \frac{P(Y = yes, X = no, Z = z)}{P(X = no|Z = z)}$$

Simpson's Paradox

X : Treatment	Y : Recovered	Z : Gender	Number	P(X,Y,Z)
Yes	Yes	Male	81	0.116
Yes	Yes	Female	192	0.274
Yes	No	Male	6	0.01
Yes	No	Female	71	0.101
No	Yes	Male	234	0.334
No	Yes	Female	55	0.079
No	No	Male	36	0.051
No	No	Female	25	0.036

Simpson's Paradox

$$P(X = yes|Z = male) = \frac{P(X=yes, Z=male)}{P(Z=male)} = \frac{0.116+0.01}{0.116+0.01+0.334+0.051} = 0.233$$

$$P(X = yes|Z = female) = \frac{P(X=yes, Z=female)}{P(Z=female)} = \frac{0.274+0.101}{0.274+0.101+0.079+0.036} = 0.765$$

$$P(X = no|Z = male) = 1 - 0.233 = 0.767$$
$$P(X = no|Z = female) = 1 - 0.765 = 0.235$$

Simpson's Paradox

$$P(Y = yes|do(X = yes)) = \sum_{z=\{male,female\}} \frac{P(Y = yes, X = yes, Z = z)}{P(X = yes|Z = z)}$$

$$= \frac{0.116}{0.233} + \frac{0.274}{0.765} = 0.498 + 0.358 = 0.856$$

$$P(Y = yes|do(X = no)) = \sum_{z=\{male,female\}} \frac{P(Y = yes, X = no, Z = z)}{P(X = no|Z = z)}$$

$$= \frac{0.335}{0.767} + \frac{0.079}{0.235} = 0.437 + 0.336 = 0.773$$

$P(Y = yes|do(X = yes))$- $P(Y = yes|do(X = no)) = \mathbf{0.856 - 0.773}$**=0.083**

Simpson's Paradox

$$P(Y = yes|do(X = yes)) = \sum_{z=\{male,female\}} \frac{P(Y = yes, X = yes, Z = z)}{P(X = yes|Z = z)}$$

$$= \frac{0.116}{0.233} + \frac{0.274}{0.765} = 0.498 + 0.358 = 0.856$$

$$P(Y = yes|do(X = no)) = \sum_{z=\{male,female\}} \frac{P(Y = yes, X = no, Z = z)}{P(X = no|Z = z)}$$

$$= \frac{0.335}{0.767} + \frac{0.079}{0.235} = 0.437 + 0.336 = 0.773$$

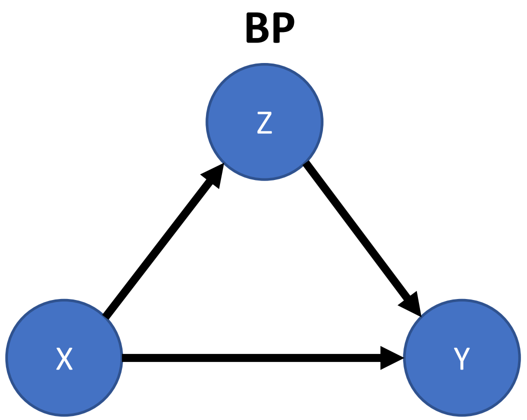
$P(Y = yes|do(X = yes))$- $P(Y = yes|do(X = no)) = \mathbf{0.856 - 0.773}$**=0.083**

Simpson's Paradox

Treatment	Low BP	High BP	Total
Yes	81/87 **(93%)**	192/263 **(73%)**	273/350 (78%)
No	234/270 (87%)	55/80 (69%)	289/350 **(83%)**

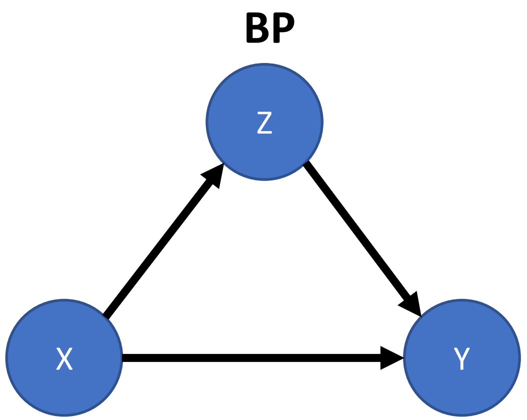
Simpson's Paradox

Treatment	Low BP	High BP	Total
Yes	81/87 **(93%)**	192/263 **(73%)**	273/350 (78%)
No	234/270 (87%)	55/80 (69%)	289/350 **(83%)**



Simpson's Paradox

Treatment	Low BP	High BP	Total
Yes	81/87 **(93%)**	192/263 **(73%)**	273/350 (78%)
No	234/270 (87%)	55/80 (69%)	289/350 **(83%)**



$$P(Y = yes|do(X = yes)) - P(Y = yes|do(X = no)) = -0.05$$

Maternal Smoking is a strong predictor of newborn mortality and low birthweight

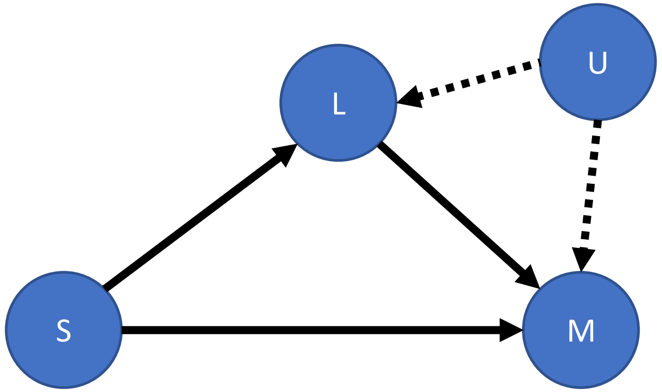
However, in newborns with low birthweight, maternal smoking is associated with lower mortality.

- Does this mean that maternal smoking is good for low birthweight newborns!?

Newborn Mortality and Maternal Smoking

Newborn Mortality and Maternal Smoking

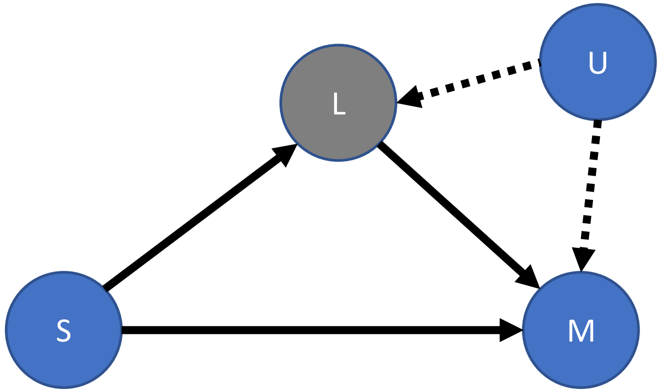
S: Maternal Smoking
M: Mortality
L: Low Birth Weight
U: Birth Defect



Sonia Hernández-Díaz, et al., The Birth Weight “Paradox” Uncovered?, *American Journal of Epidemiology*, Volume 164, Issue 11, 1 December 2006, Pages 1115–1120

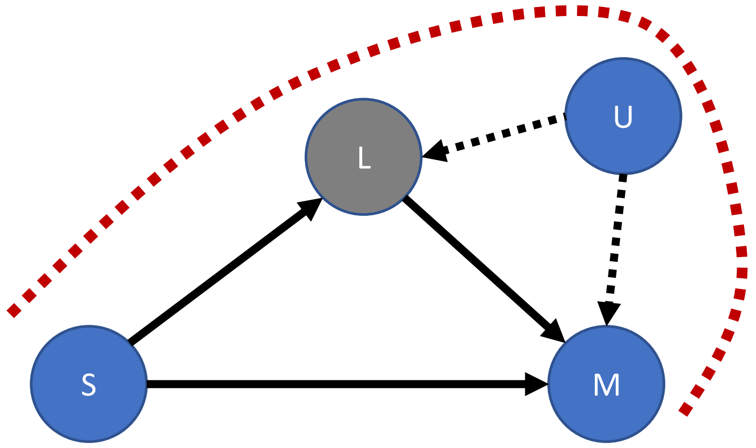
Newborn Mortality and Maternal Smoking

S: Maternal Smoking
M: Mortality
L: Low Birth Weight
U: Birth Defect



Newborn Mortality and Maternal Smoking

S: Maternal Smoking
M: Mortality
L: Low Birth Weight
U: Birth Defect



- A group of 47 editors of 35 respiratory, sleep, and critical care journals
  - They urge authors to consider using causal models (DAGs)

PERSPECTIVE

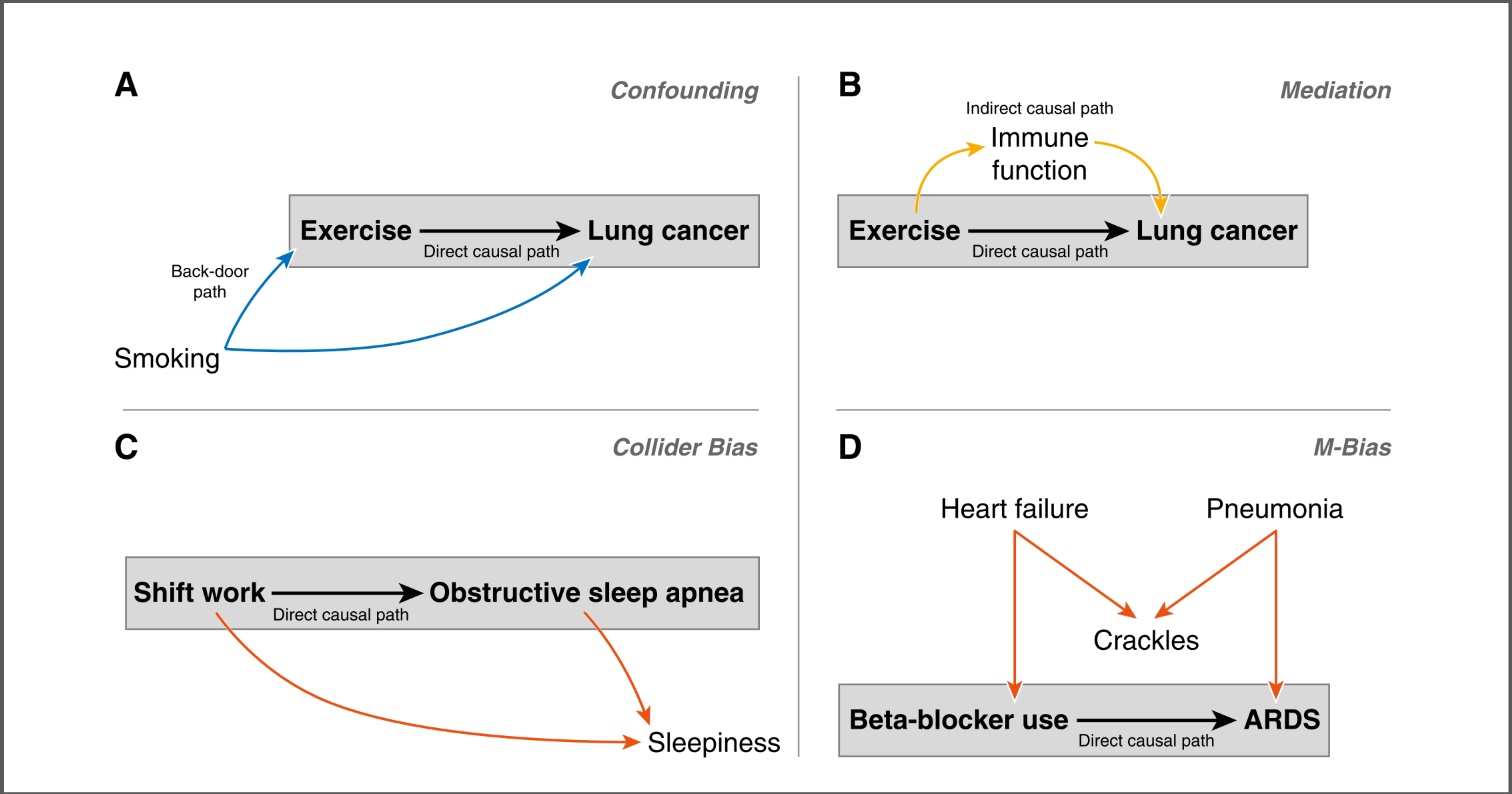
SPECIAL SECTION

Control of Confounding and Reporting of Results in Causal Inference Studies

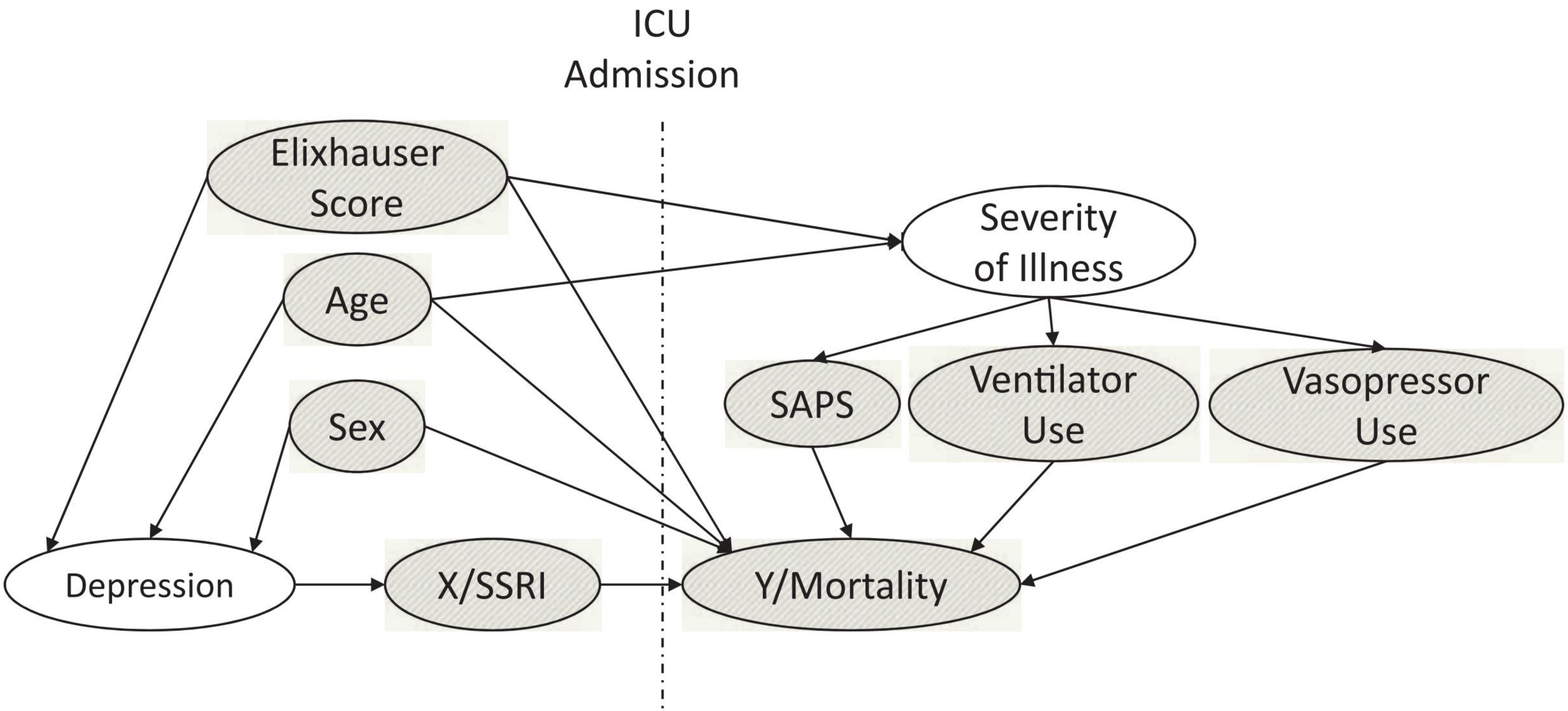
Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals

David J. Lederer[1,2]*, Scott C. Bell[3]*, Richard D. Branson[4]*, James D. Chalmers[5]*, Rachel Marshall[6]*, David M. Maslove[7]*, David E. Ost[8]*, Naresh M. Punjabi[9]*, Michael Schatz[10]*, Alan R. Smyth[11]*, Paul W. Stewart[12]*, Samy Suissa[13]*, Alex A. Adjei[14], Cezmi A. Akdis[15], Élie Azoulay[16], Jan Bakker[17,18,19], Zuhair K. Ballas[20], Philip G. Bardin[21], Esther Barreiro[22], Rinaldo Bellomo[23], Jonathan A. Bernstein[24], Vito Brusasco[25], Timothy G. Buchman[26,27,28], Sudhansu Chokroverty[29], Nancy A. Collop[30,31], James D. Crapo[32], Dominic A. Fitzgerald[33], Lauren Hale[34], Nicholas Hart[35], Felix J. Herth[36], Theodore J. Iwashyna[37], Gisli Jenkins[38], Martin Kolb[39], Guy B. Marks[40], Peter Mazzone[41], J. Randall Moorman[42,43,44], Thomas M. Murphy[45], Terry L. Noah[46], Paul Reynolds[47], Dieter Riemann[48], Richard E. Russell[49,50], Aziz Sheikh[51], Giovanni Sotgiu[52], Erik R. Swenson[53], Rhonda Szczesniak[54,55], Ronald Szymusiak[56,57], Jean-Louis Teboul[58], and Jean-Louis Vincent[59]

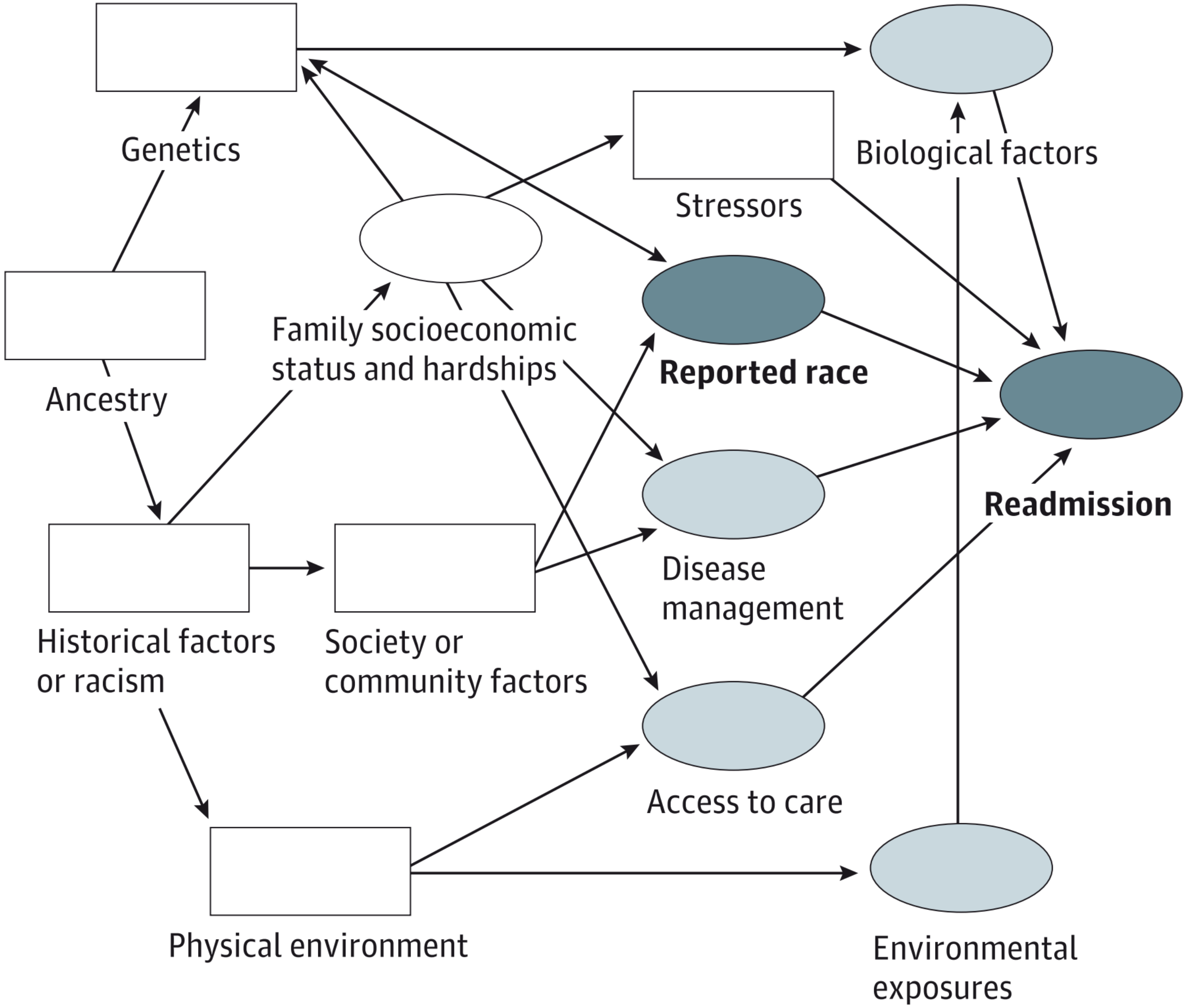
Lederer, David J. et al. (2019): Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. In *Annals of the American Thoracic Society*



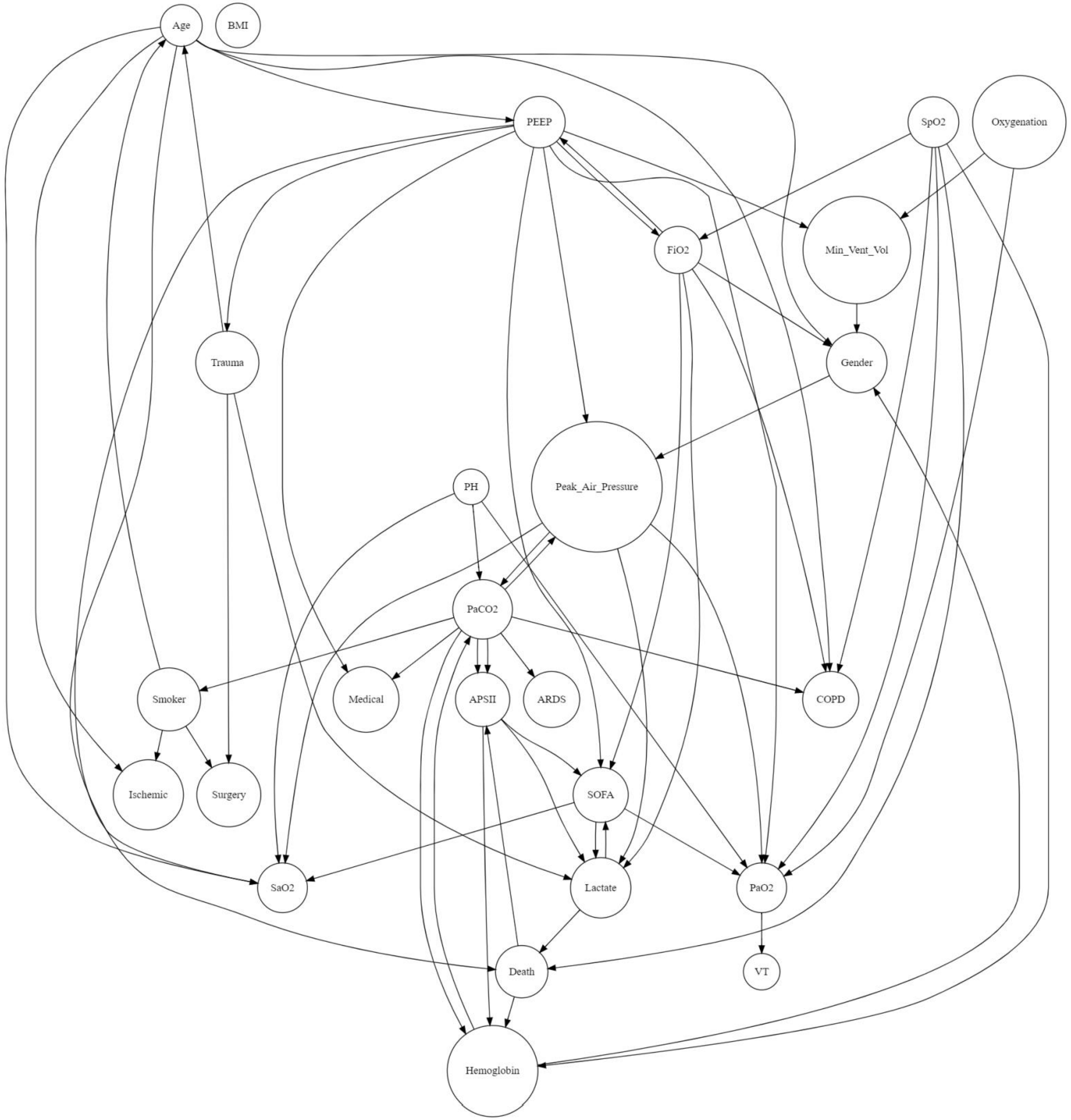
Lederer, David J. et al. (2019): Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. In *Annals of the American Thoracic Society*



Ghassemi, M. et al., 2014. Leveraging a critical care database: selective serotonin reuptake inhibitor use prior to ICU admission is associated with increased hospital mortality. Chest, 145(4), pp.745-752.



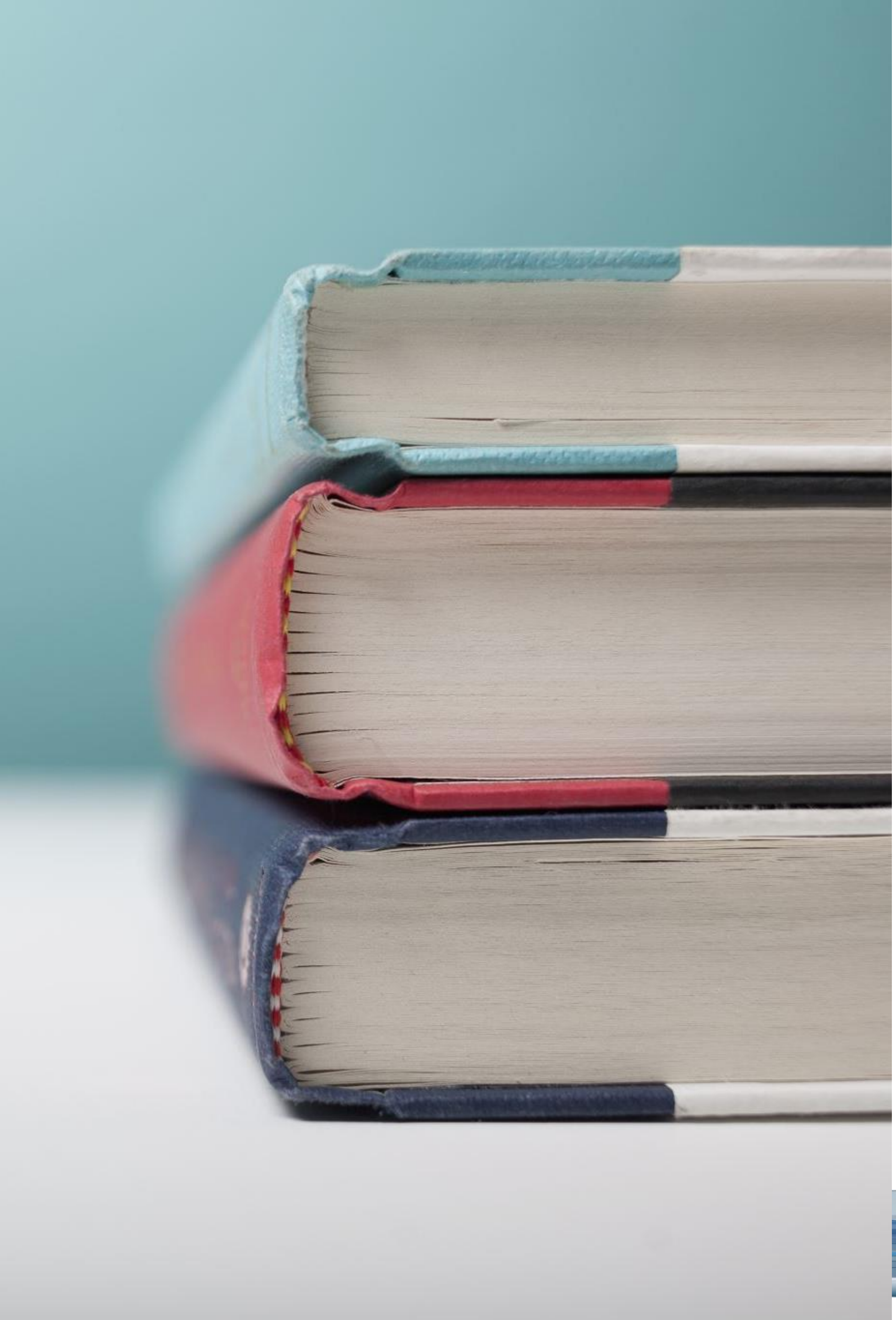
Beck, A.F. et al., 2016. Explaining racial disparities in child asthma readmission using a causal inference approach. *JAMA pediatrics*, *170*(7), pp.695-703.



Gani, M.O., et al. 2020. Structural Causal Model with Expert Augmented Knowledge to Estimate the Effect of Oxygen Therapy on Mortality in the ICU. *arXiv preprint*

Other Topics

- Propensity Score
- Counterfactual
- Difference-in-differences
- Regression Discontinuity
- Instrumental variables
- G-Estimation



Reading Suggestions

- Gaskell, Amy L.; Sleigh, Jamie W. (2020): An Introduction to Causal Diagrams for Anesthesiology Research. In *Anesthesiology* 132 (5), pp. 951–967.
- Pearl, Judea; Mackenzie, Dana (2018): The book of why. The new science of cause and effect. New York: Basic Books.
- Pearl, Judea; Glymour, Madelyn; Jewell, Nicholas P. (2016): Causal inference in statistics. A primer / Judea Pearl, Madelyn Glymour, Nicholas Jewell. 1st. Hoboken, New Jersey: John Wiley & Sons.
- Rosenbaum, Paul R. (2017): Observation and Experiment. An introduction to causal inference / Paul R. Rosenbaum. Cambridge, Massachusetts: Harvard University Press.
- Gelman, Andrew; Hill, Jennifer; Vehtari, Aki (2021): Regression and other stories. Cambridge: Cambridge University Press (Analytical methods for social research).
- Miguel A Hernan; James M. Robins (2020): Casual Inference. What If.

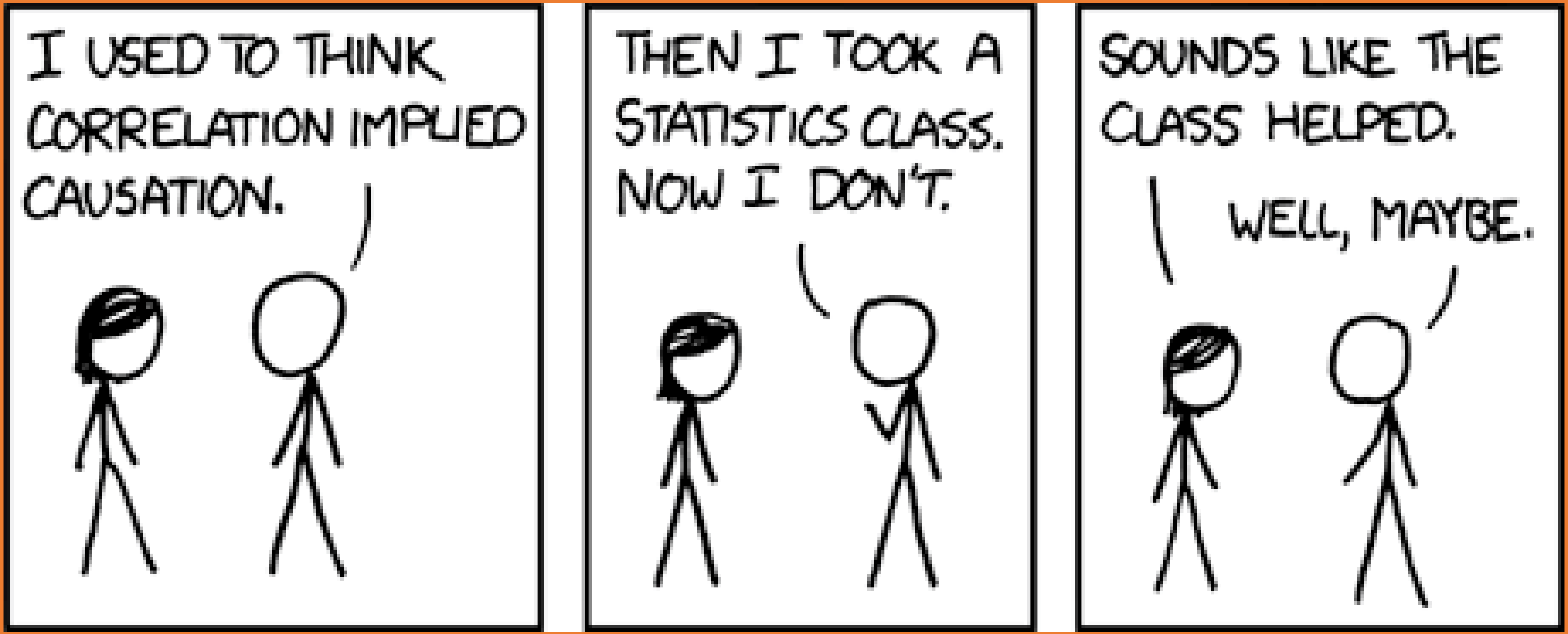
Summary

Controlling for all covariates are generally wrong.

With expert knowledge, we can model data generation process using DAGs.

Using DAGs

- Check our assumptions
- Identify causal effect from observational data



https://xkcd.com/552/

Thank you!