Causal Discovery from Observational Data

Brady Neal

causalcourse.com

What if we don’t have the causal graph?

What if we don't have the causal graph?

Causal discovery: data ⟶ causal graph

What if we don't have the causal graph?

Causal discovery: data $\longrightarrow$ causal graph

Structure identification: identifying the causal graph

Independence-Based Causal Discovery

Semi-Parametric Causal Discovery

Faithfulness Assumption

Faithfulness Assumption

Recall the Markov assumption: $X \perp\!\!\!\perp_G Y \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z$

Faithfulness Assumption

Recall the Markov assumption: $X \perp\!\!\!\perp_G Y \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z$

Causal graph → Data

Faithfulness Assumption

Recall the Markov assumption: $X \perp\!\!\!\perp_G Y \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z$

Causal graph → Data

Causal graph ← Data

Faithfulness Assumption

Recall the Markov assumption: $X \perp\!\!\!\perp_G Y \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z$

Causal graph → Data

Causal graph ← Data

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \impliedby X \perp\!\!\!\perp_P Y \mid Z$

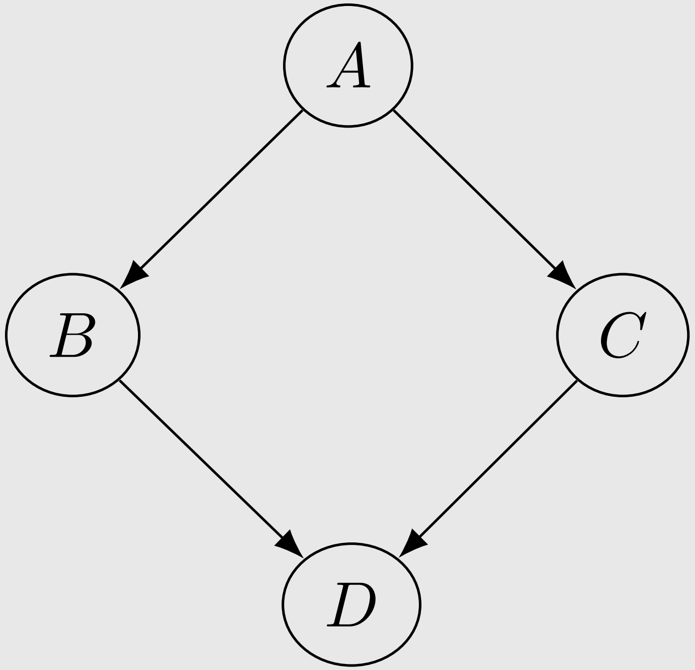
Violation of Faithfulness

Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \impliedby X \perp\!\!\!\perp_P Y \mid Z$

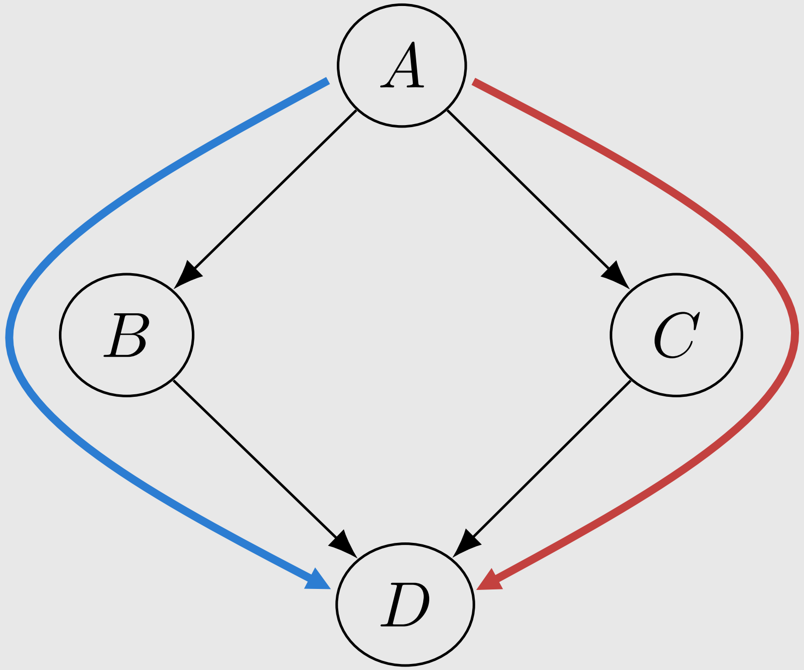
Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \impliedby X \perp\!\!\!\perp_P Y \mid Z$



Violation of Faithfulness

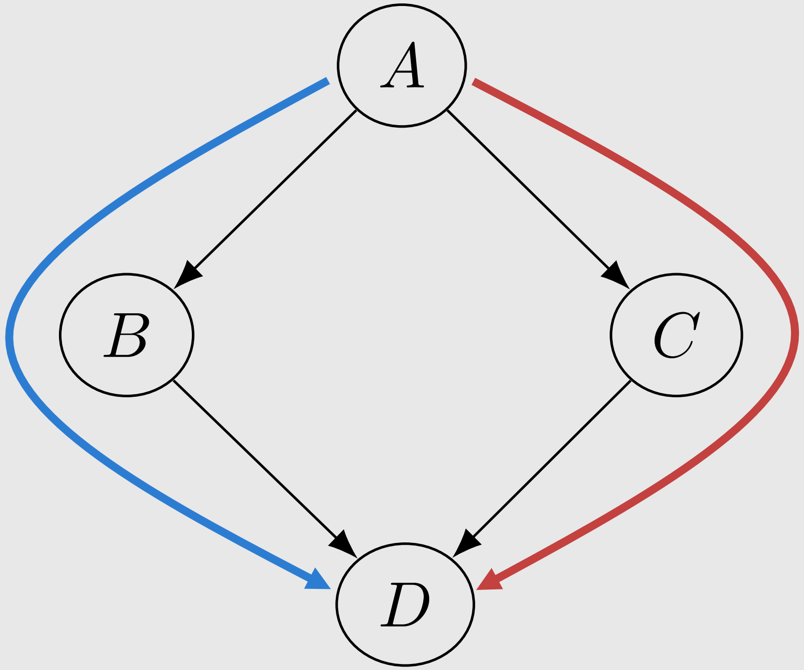
Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$



Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

$A \perp\!\!\!\perp D$

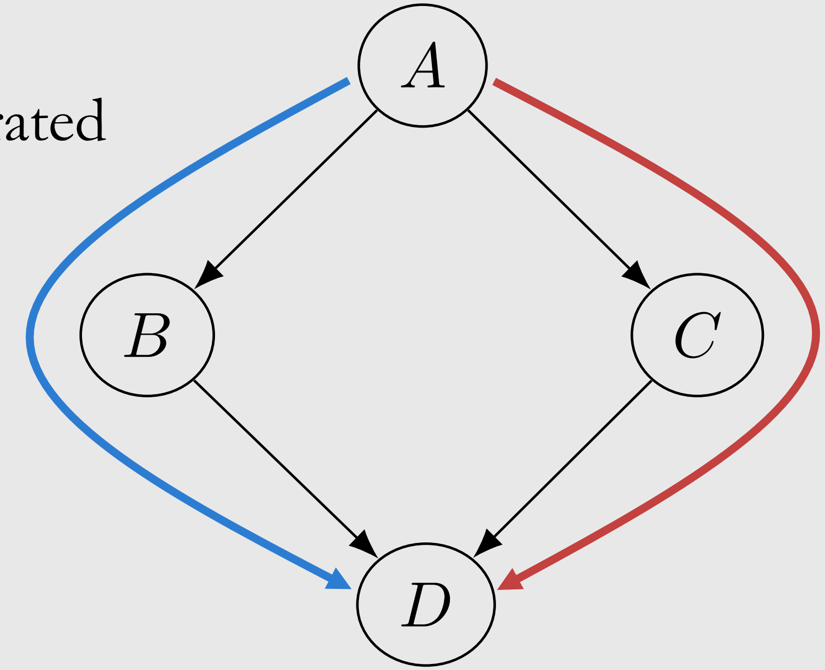


Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

$A \perp\!\!\!\perp D$

but A and D aren't d-separated

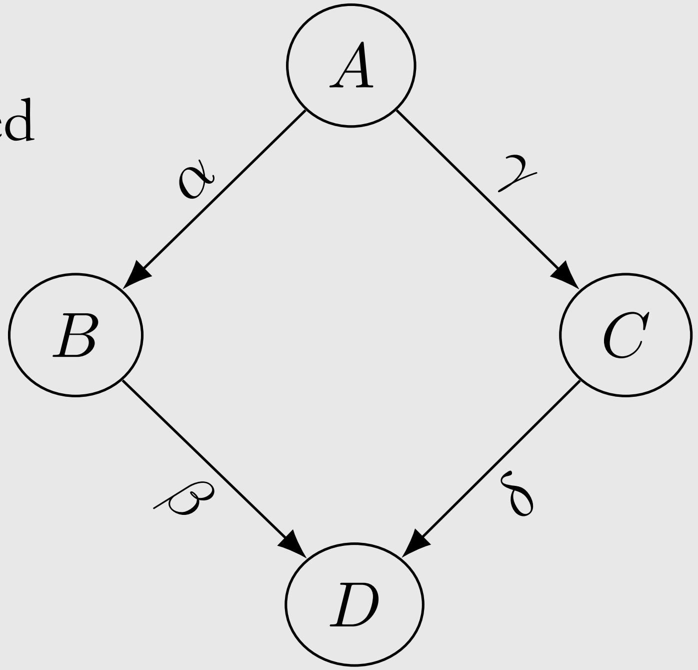


Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

$A \perp\!\!\!\perp D$

but A and D aren't d-separated

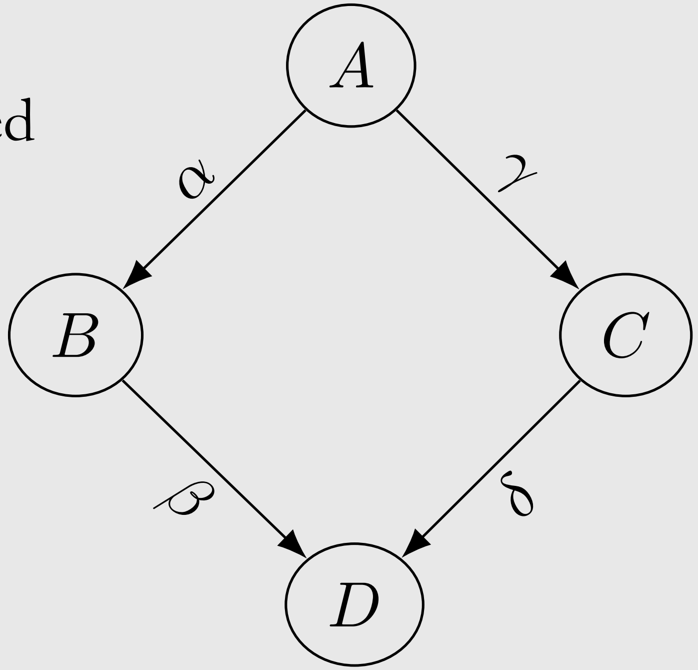


Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

$A \perp\!\!\!\perp D$

but A and D aren't d-separated



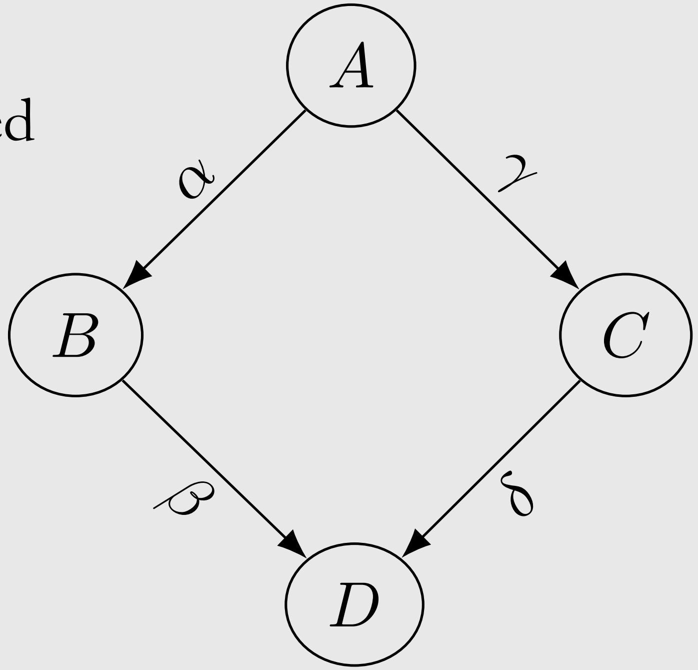
$$B := \alpha A$$
$$C := \gamma A$$
$$D := \beta B + \delta C$$

Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

$$A \perp\!\!\!\perp D$$

but A and D aren't d-separated



$B := \alpha A$

$C := \gamma A$

$D := \beta B + \delta C$

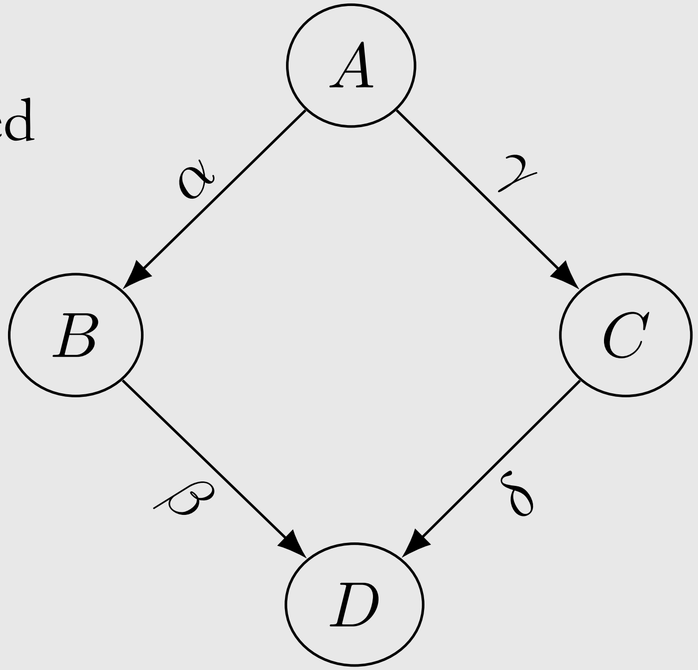
$$D = (\alpha\beta + \gamma\delta)A$$

Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

$A \perp\!\!\!\perp D$

but A and D aren't d-separated



$$B := \alpha A$$
$$C := \gamma A$$
$$D := \beta B + \delta C$$

$$D = (\alpha\beta + \gamma\delta) A$$

Paths cancel if $\alpha\beta = -\gamma\delta$

Causal Sufficiency and Acyclicity

Causal Sufficiency and Acyclicity

Causal Sufficiency: there are no unobserved confounders of any of the variables in the graph.

Causal Sufficiency and Acyclicity

Causal Sufficiency: there are no unobserved confounders of any of the variables in the graph.

Acyclicity: still assuming there are no cycles in the graph.

Causal Sufficiency and Acyclicity

Causal Sufficiency: there are no unobserved confounders of any of the variables in the graph.

Acyclicity: still assuming there are no cycles in the graph.

All assumptions:

- Markov assumption
- Faithfulness
- Causal sufficiency
- Acyclicity

Causal Sufficiency and Acyclicity

Causal Sufficiency: there are no unobserved confounders of any of the variables in the graph.

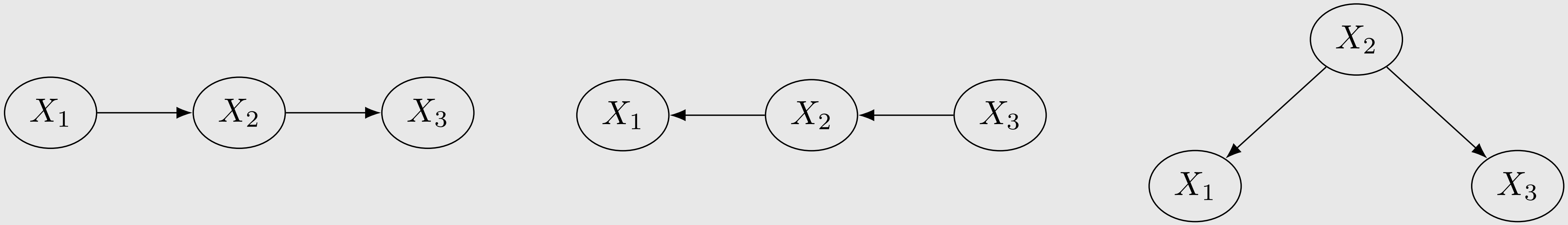
Acyclicity: still assuming there are no cycles in the graph.

All assumptions:

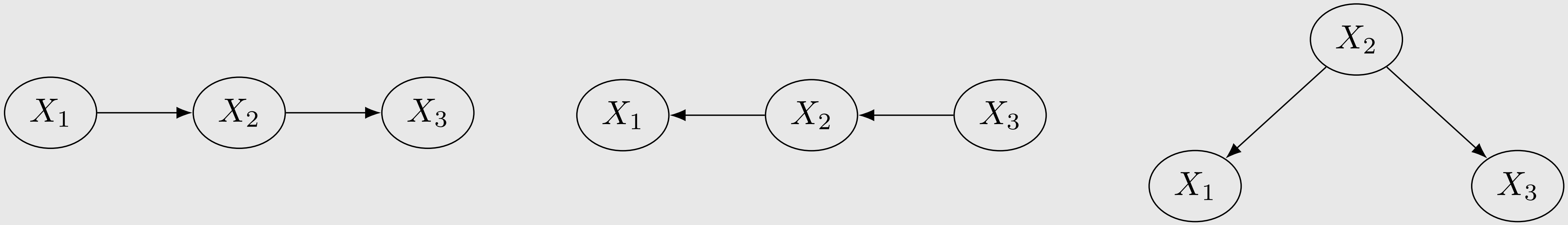
- Markov assumption
- Faithfulness
- Causal sufficiency
- Acyclicity

Question:
Why is the Markov assumption (plus causal sufficiency and acyclicity) not enough for learning causal graphs from data?

Chains and Forks Encode Same Independencies

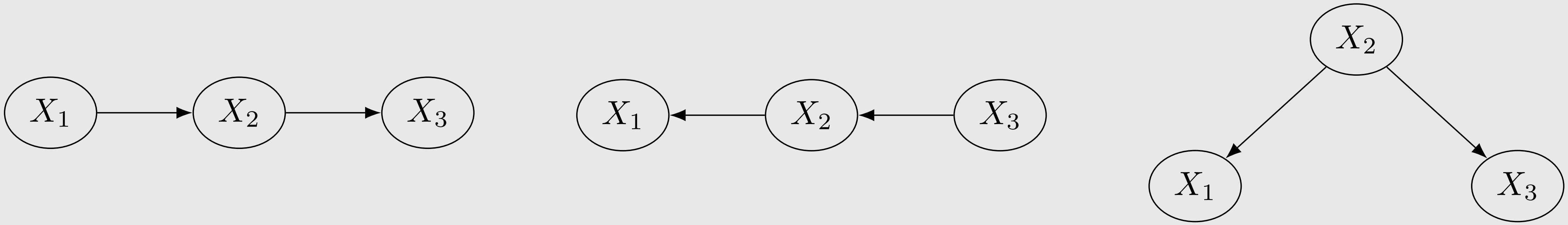


Chains and Forks Encode Same Independencies



Markov: $X_1 \perp\!\!\!\perp X_3 \mid X_2$

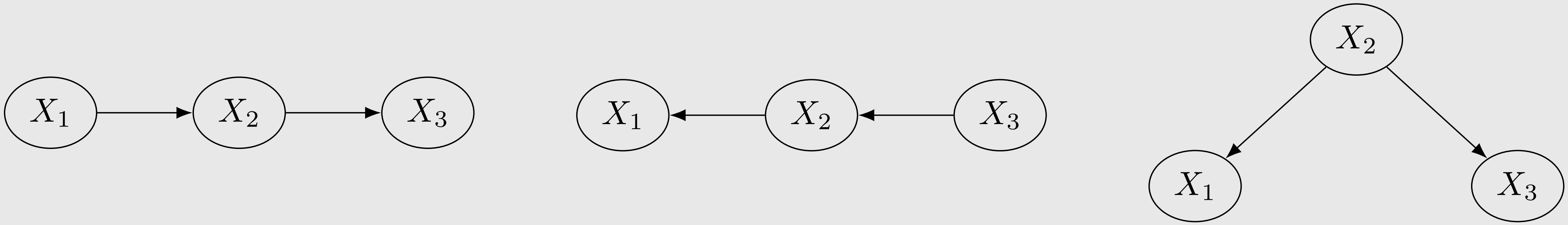
Chains and Forks Encode Same Independencies



Markov: $X_1 \perp\!\!\!\perp X_3 \mid X_2$

Minimality: $X_1 \not\perp\!\!\!\perp X_2$ and $X_2 \not\perp\!\!\!\perp X_3$

Chains and Forks Encode Same Independencies



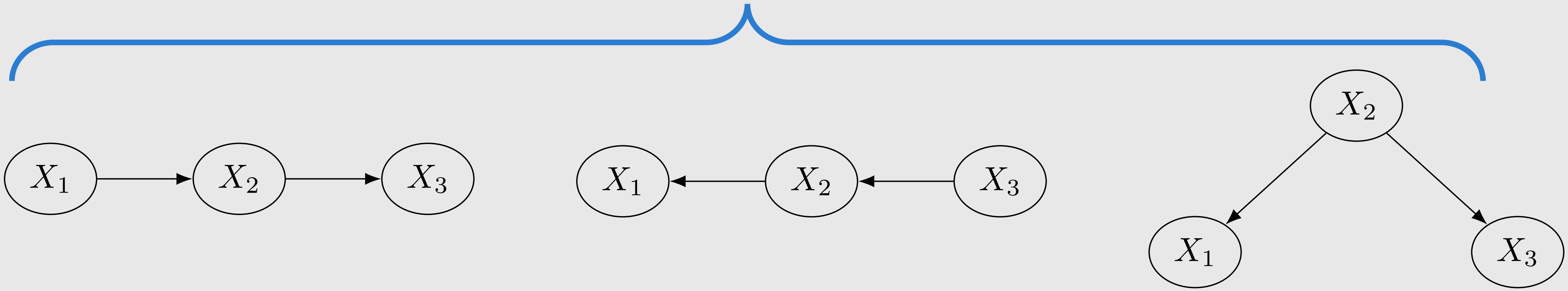
Markov: $X_1 \perp\!\!\!\perp X_3 \mid X_2$

Minimality: $X_1 \not\perp\!\!\!\perp X_2$ and $X_2 \not\perp\!\!\!\perp X_3$

Faithfulness: $X_1 \not\perp\!\!\!\perp X_3$

Chains and Forks Encode Same Independencies

Markov equivalent (all in the same Markov equivalence class)



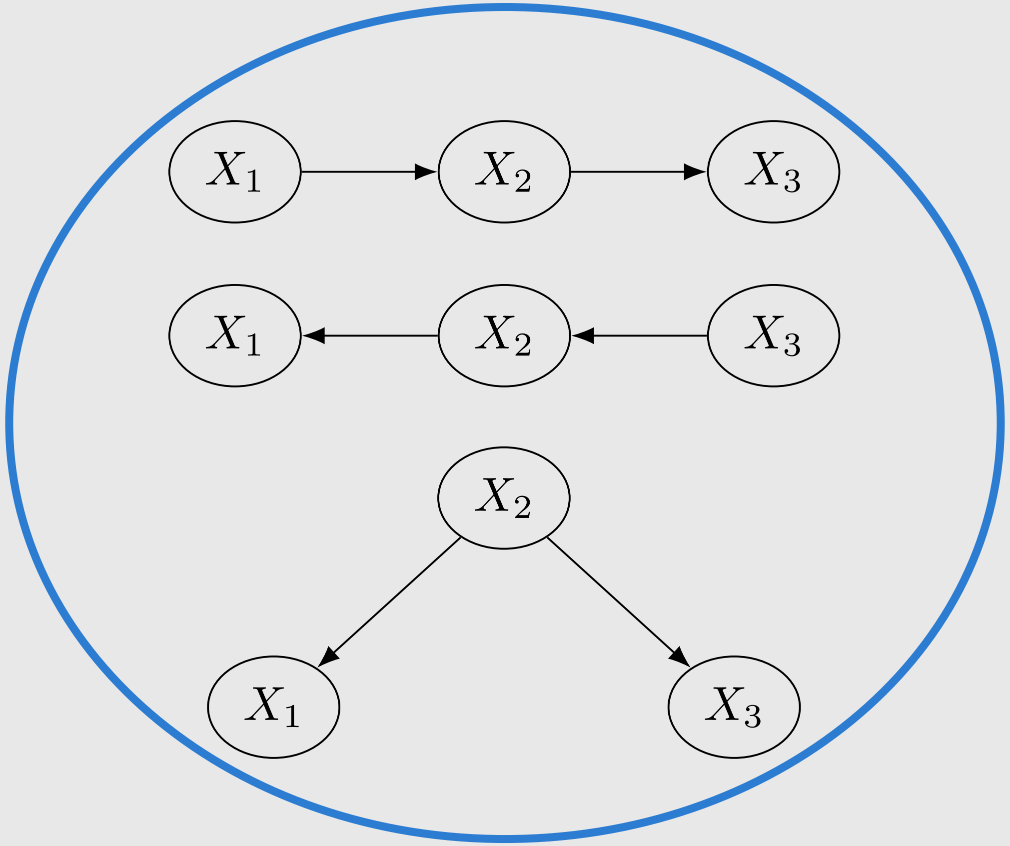
Markov: $X_1 \perp\!\!\!\perp X_3 \mid X_2$

Minimality: $X_1 \not\perp\!\!\!\perp X_2$ and $X_2 \not\perp\!\!\!\perp X_3$

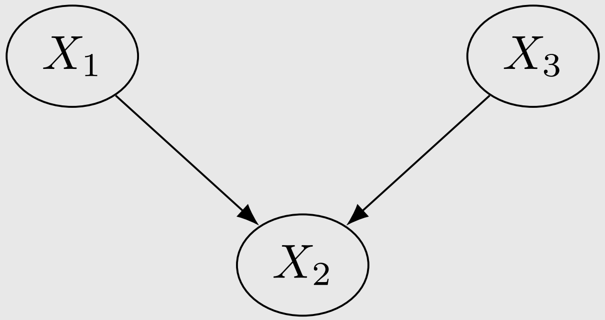
Faithfulness: $X_1 \not\perp\!\!\!\perp X_3$

Immoralities are Special

Markov equivalence class where
$X_1 \perp\!\!\!\perp X_3 \mid X_2$ and $X_1 \not\perp\!\!\!\perp X_3$

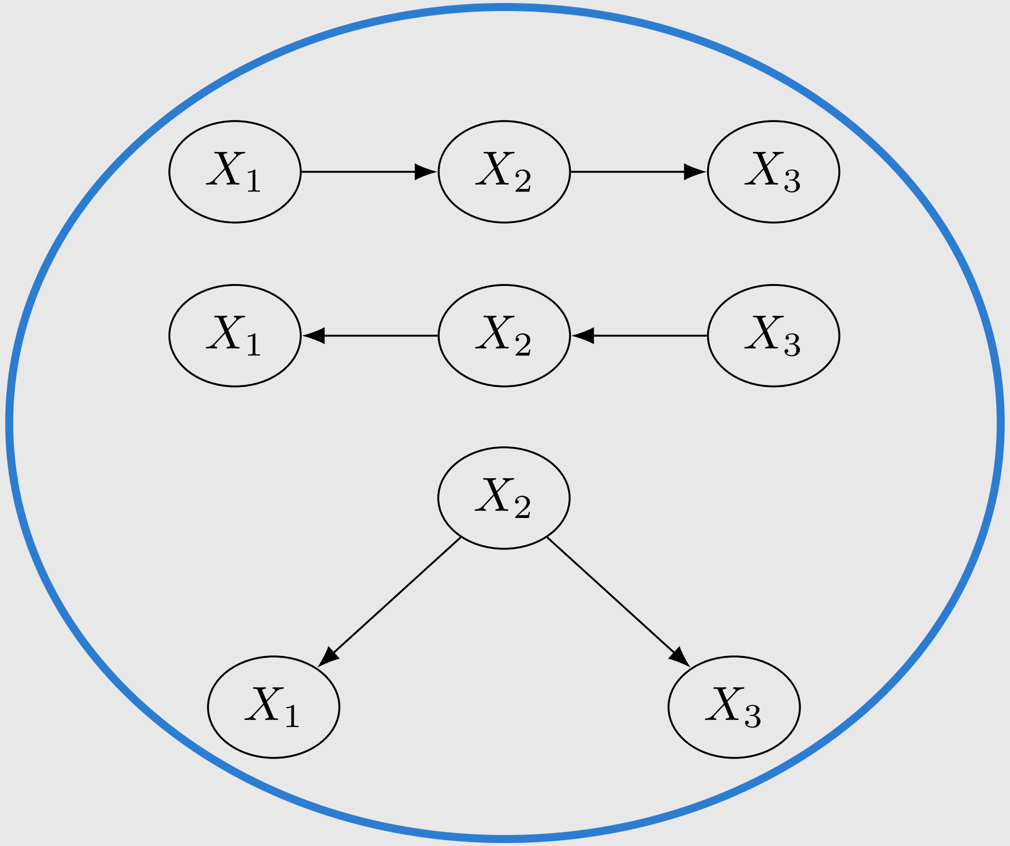


Immoralities are Special



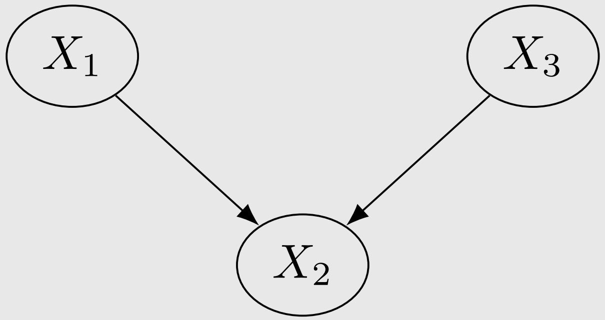
Markov equivalence class where

$X_1 \perp\!\!\!\perp X_3 \mid X_2$ and $X_1 \not\perp\!\!\!\perp X_3$

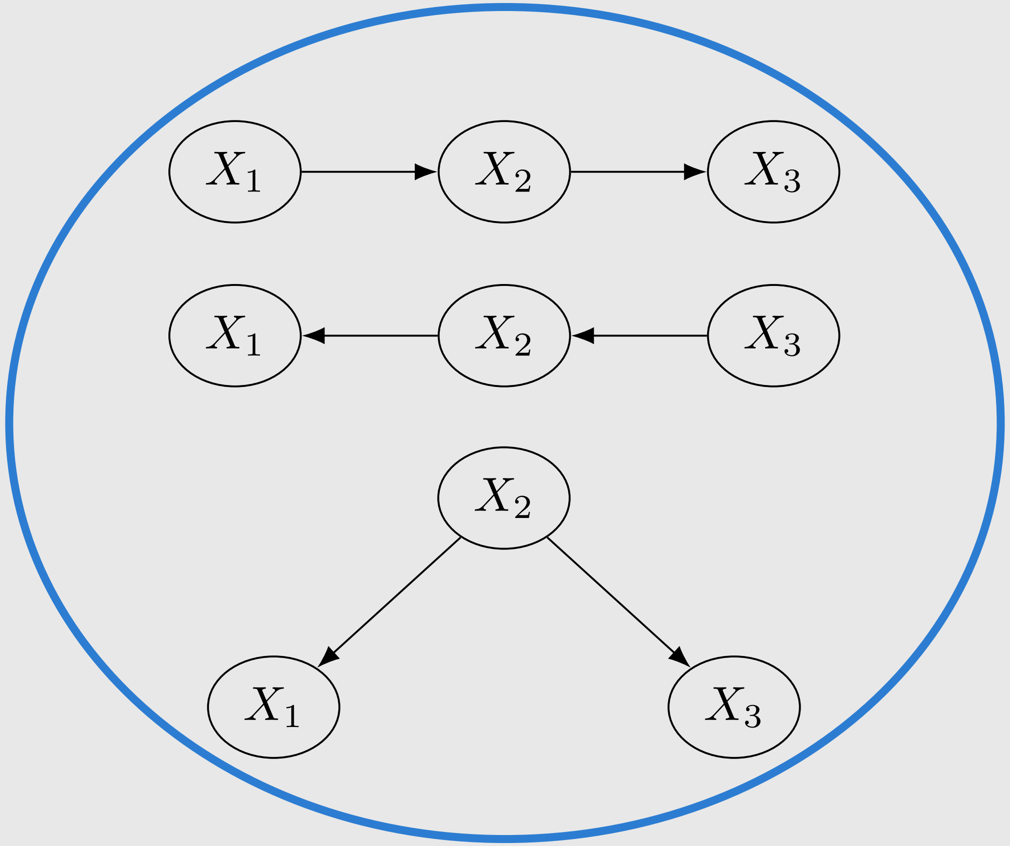


Immoralities are Special

$X_1 \perp\!\!\!\perp X_3$



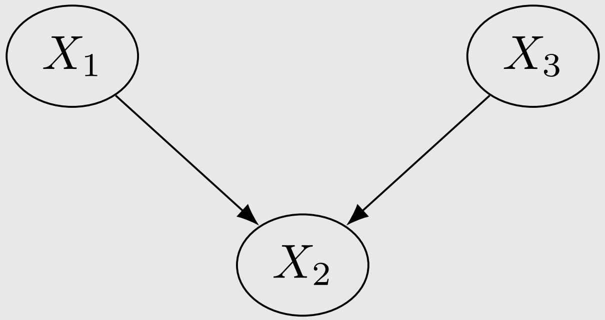
Markov equivalence class where
$X_1 \perp\!\!\!\perp X_3 \mid X_2$ and $X_1 \not\perp\!\!\!\perp X_3$

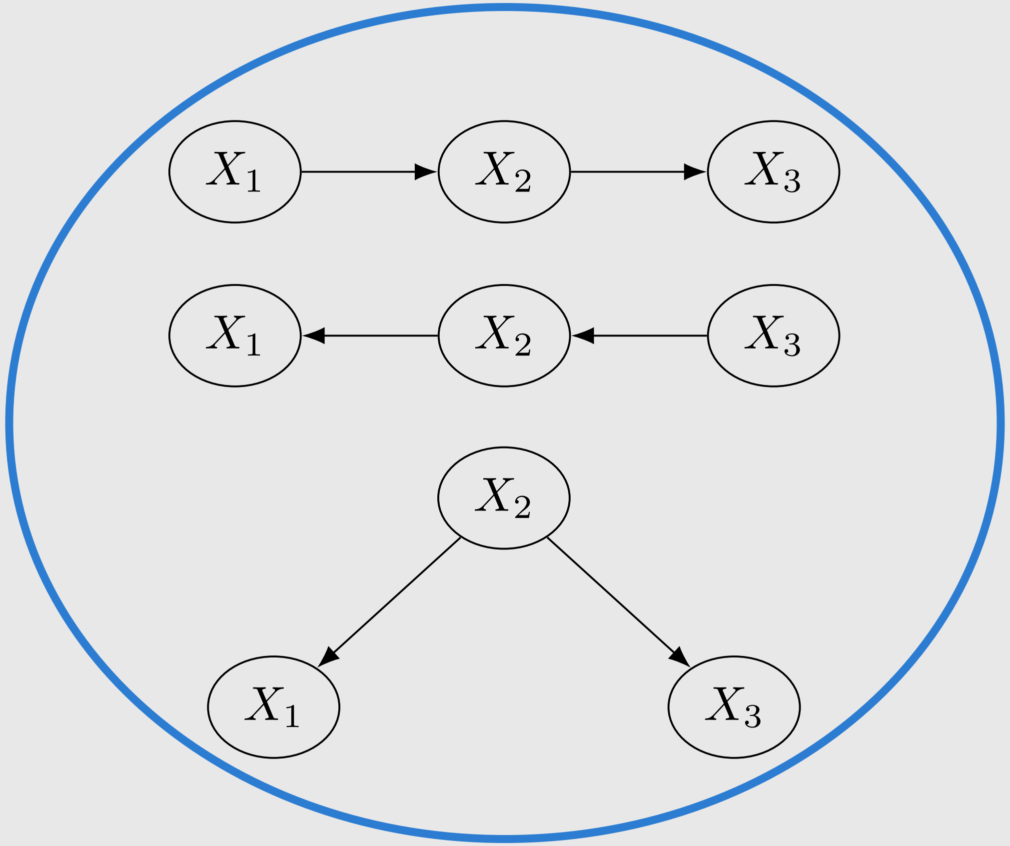


Immoralities are Special

Markov equivalence class where
$X_1 \perp\!\!\!\perp X_3 \mid X_2$ and $X_1 \not\!\perp\!\!\!\perp X_3$

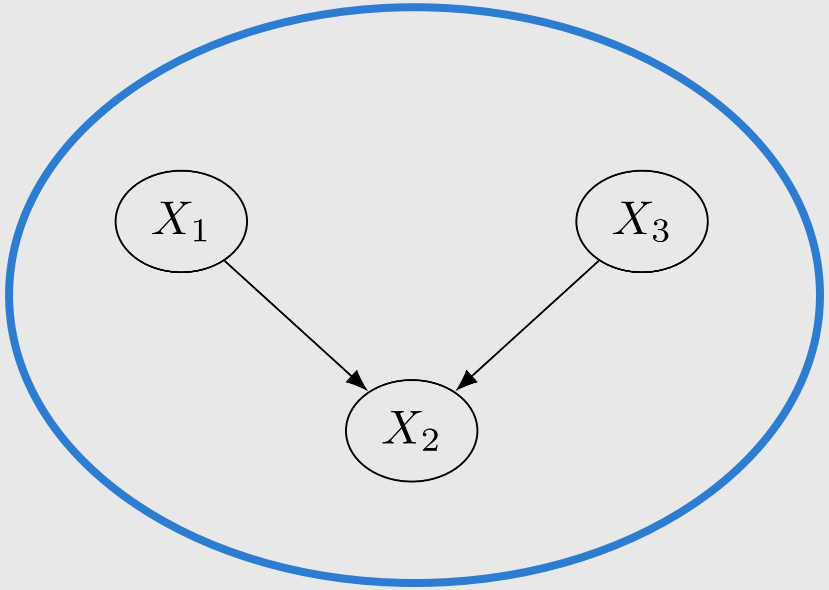
$X_1 \perp\!\!\!\perp X_3$ and $X_1 \not\!\perp\!\!\!\perp X_3 \mid X_2$



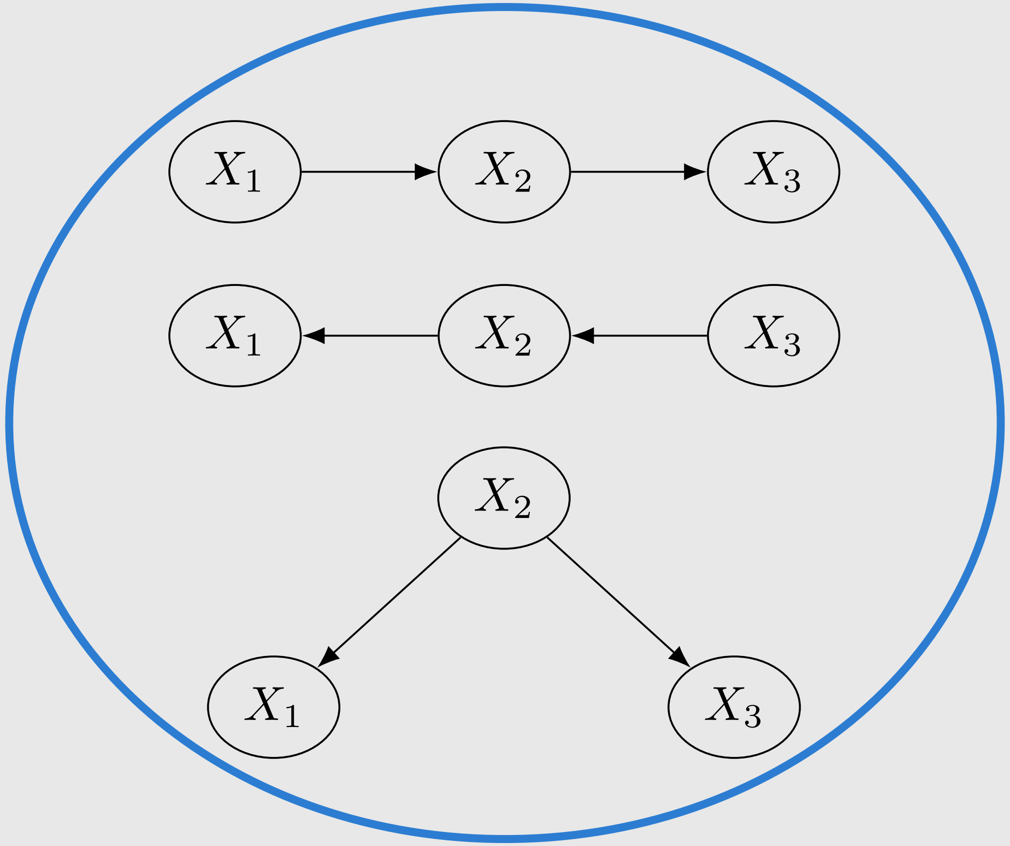


Immoralities are Special

Markov equivalence class where
$X_1 \perp\!\!\!\perp X_3$ and $X_1 \not\perp\!\!\!\perp X_3 \mid X_2$

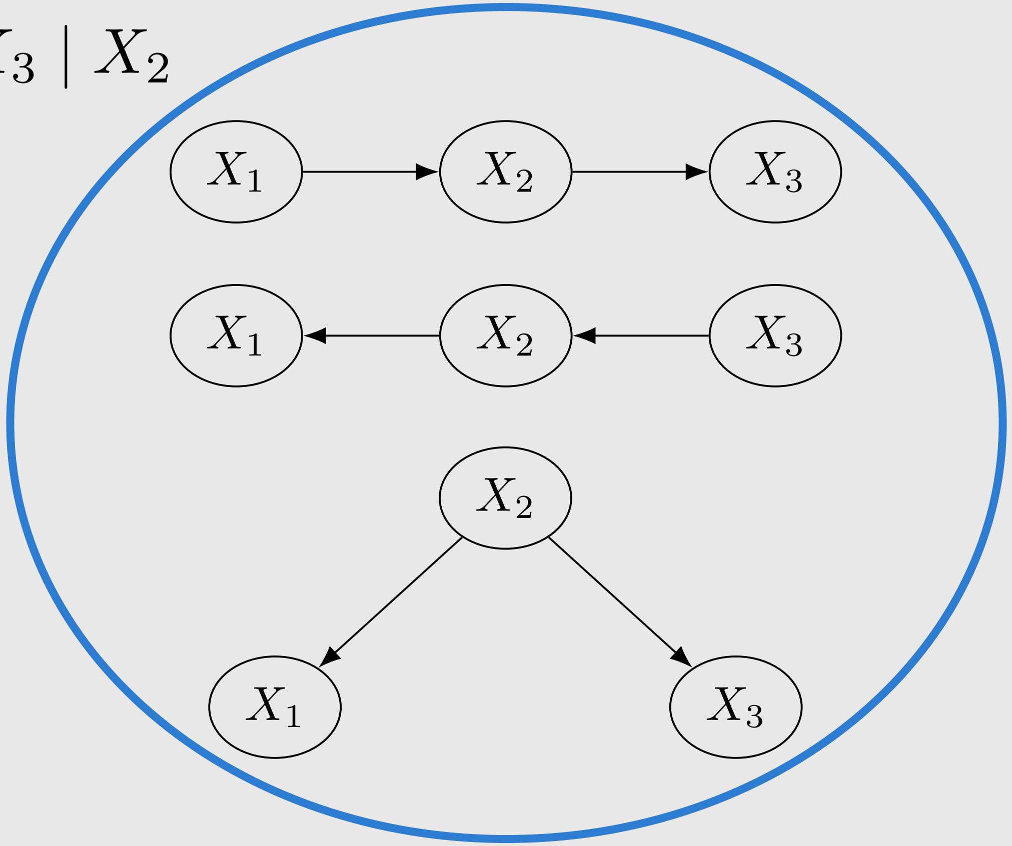


Markov equivalence class where
$X_1 \perp\!\!\!\perp X_3 \mid X_2$ and $X_1 \not\perp\!\!\!\perp X_3$



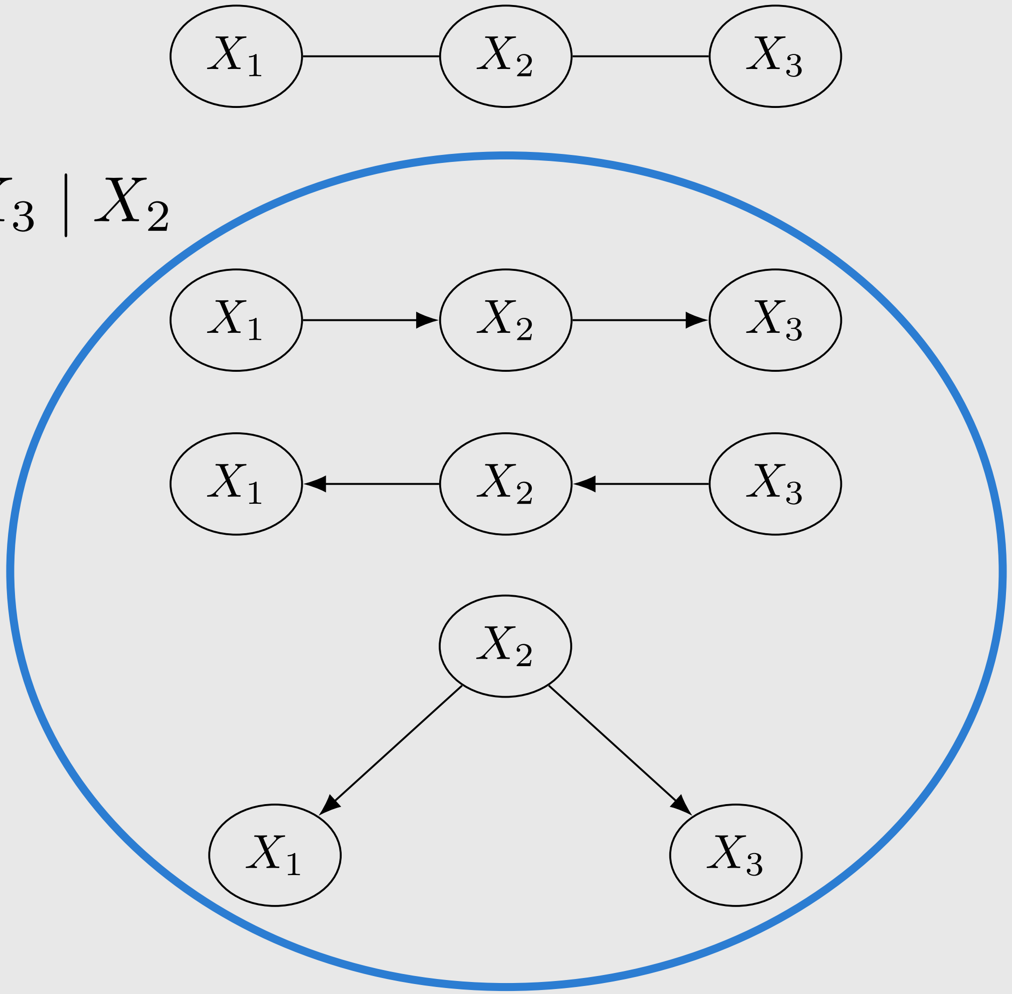
Skeletons

$X_1 \perp\!\!\!\perp X_3 \mid X_2$

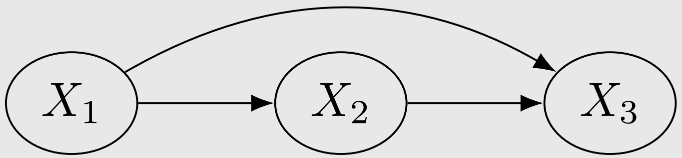


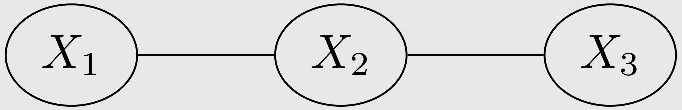
Skeletons

$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$

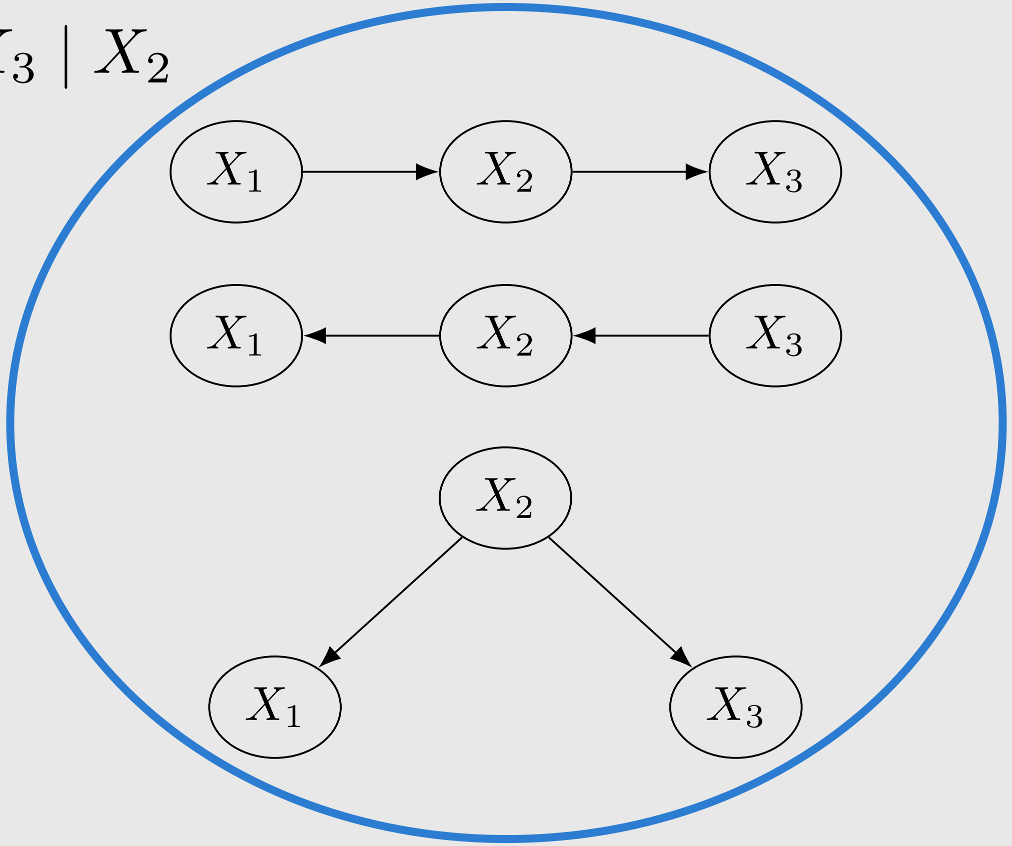


Skeletons



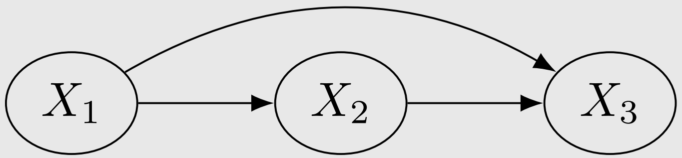


$X_1 \perp\!\!\!\perp X_3 \mid X_2$

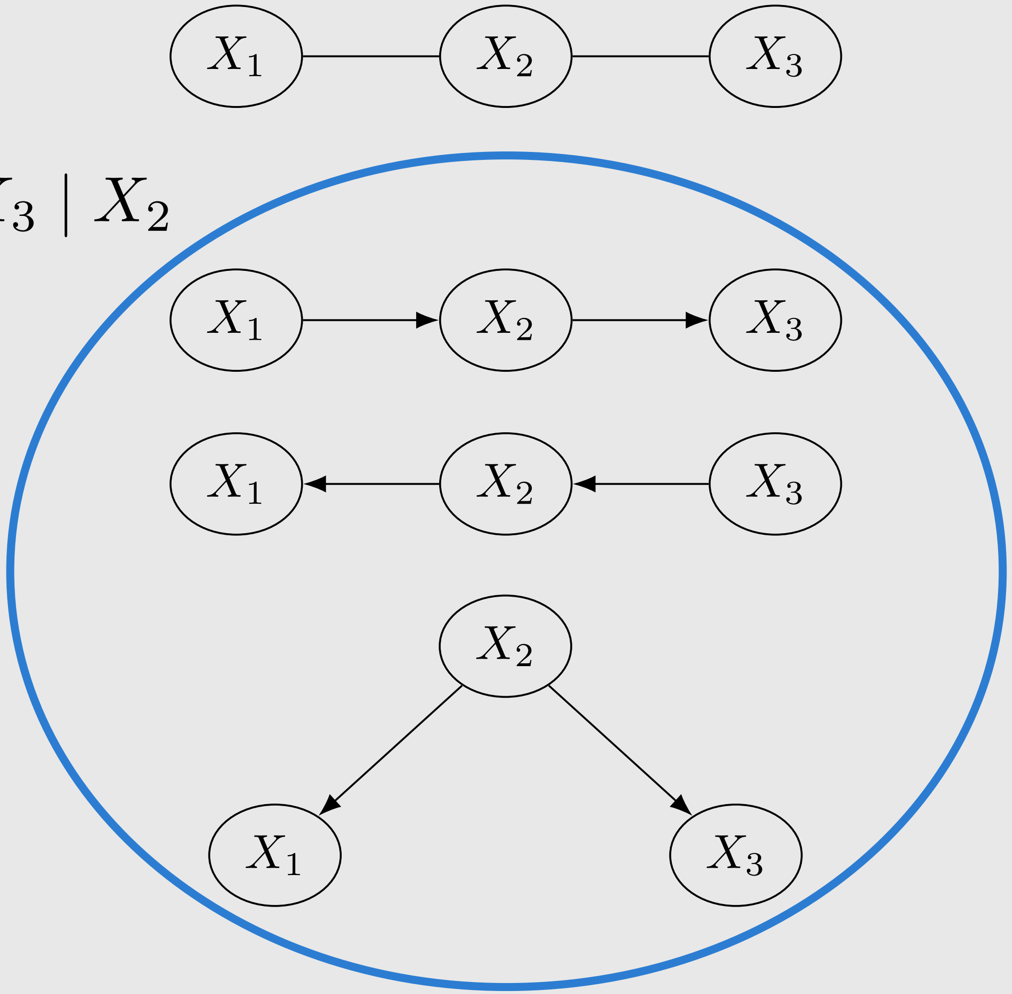


Skeletons

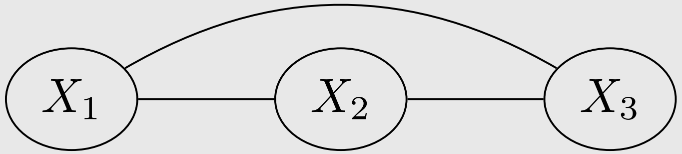
$X_1 \not\perp\!\!\!\perp X_3 \mid X_2$



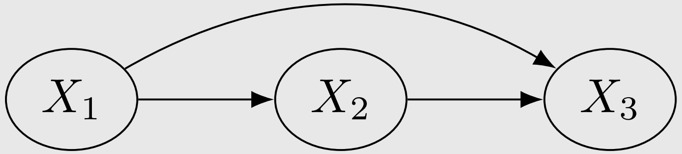
$X_1 \perp\!\!\!\perp X_3 \mid X_2$

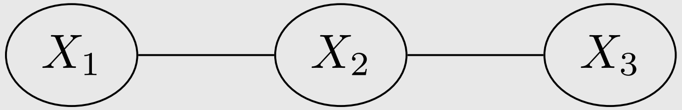


Skeletons

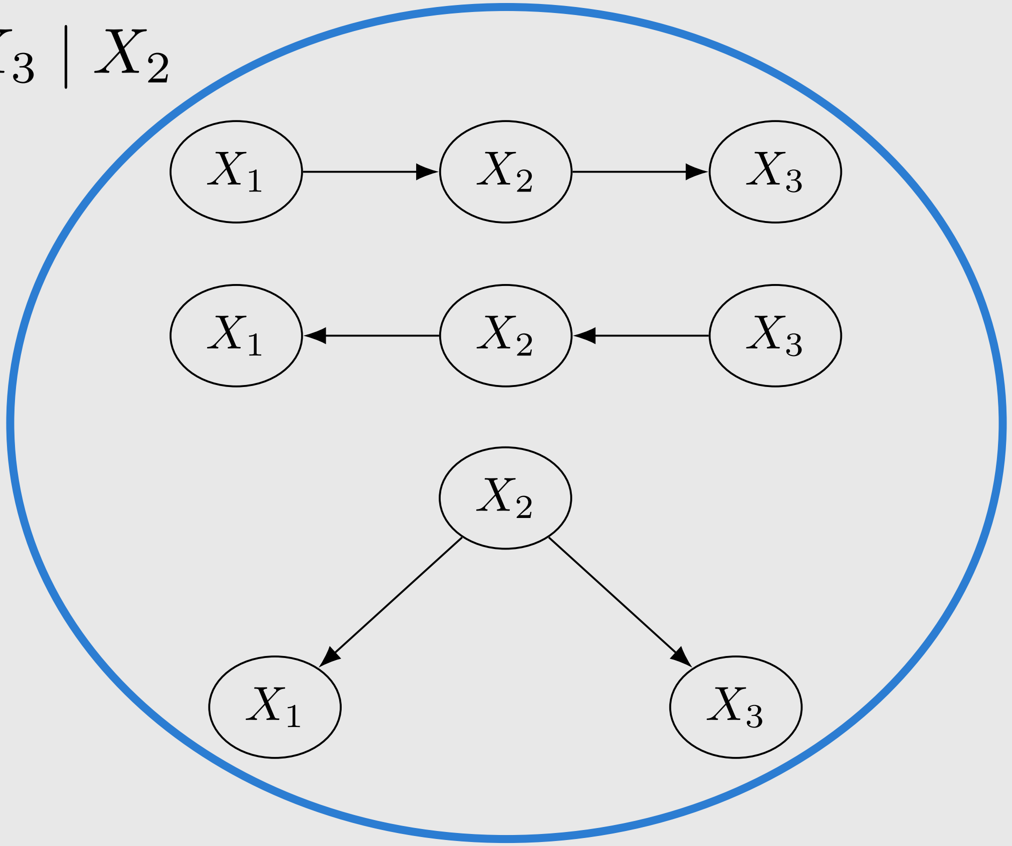


$$X_1 \not\perp\!\!\!\perp X_3 \mid X_2$$





$$X_1 \perp\!\!\!\perp X_3 \mid X_2$$



Markov Equivalence via Immoral Skeletons

Markov Equivalence via Immoral Skeletons

Two important graph qualities that we can use to distinguish graphs:

Markov Equivalence via Immoral Skeletons

Two important graph qualities that we can use to distinguish graphs:

1. Immoralities

Markov Equivalence via Immoral Skeletons

Two important graph qualities that we can use to distinguish graphs:

1. Immoralities
2. Skeleton

Markov Equivalence via Immoral Skeletons

Two important graph qualities that we can use to distinguish graphs:

1. Immoralities
2. Skeleton

Theorem: Two graphs are Markov equivalent if and only if they have the same skeleton and same immoralities (Verma & Pearl, 1990; Frydenburg, 1990).

Markov Equivalence via Immoral Skeletons

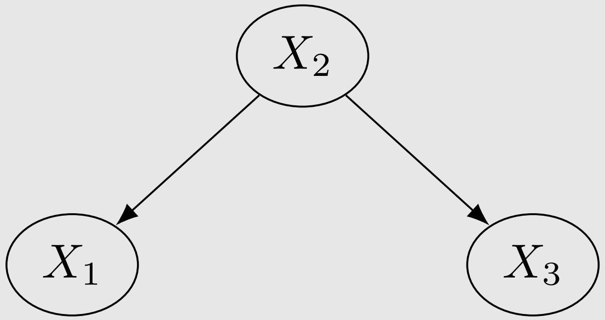
Two important graph qualities that we can use to distinguish graphs:

1. Immoralities
2. Skeleton

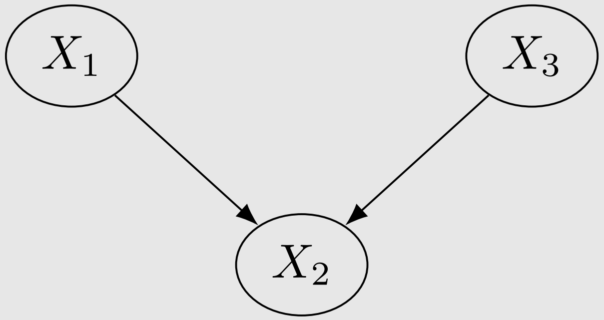
Theorem: Two graphs are Markov equivalent if and only if they have the same skeleton and same immoralities (Verma & Pearl, 1990; Frydenburg, 1990).

Essential graph (aka CPDAG): skeleton + immoralities

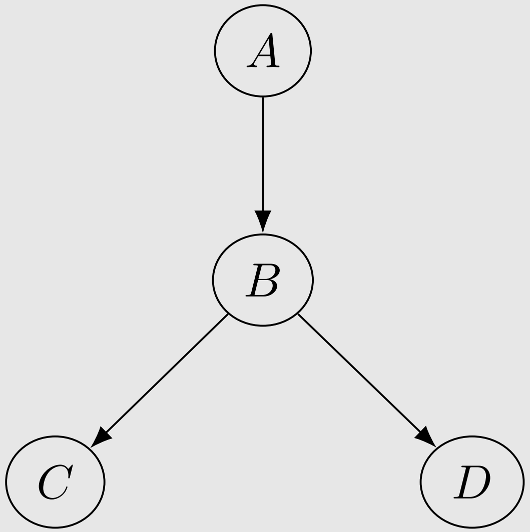
Question:
What graphs are Markov equivalent to the basic fork graph?



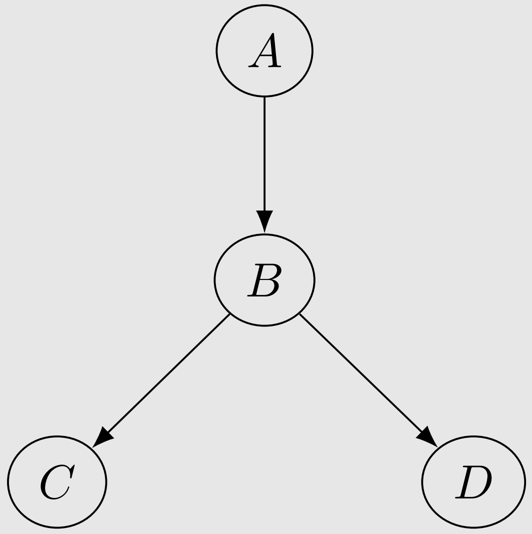
Question:
What graphs are Markov equivalent to the basic immorality?



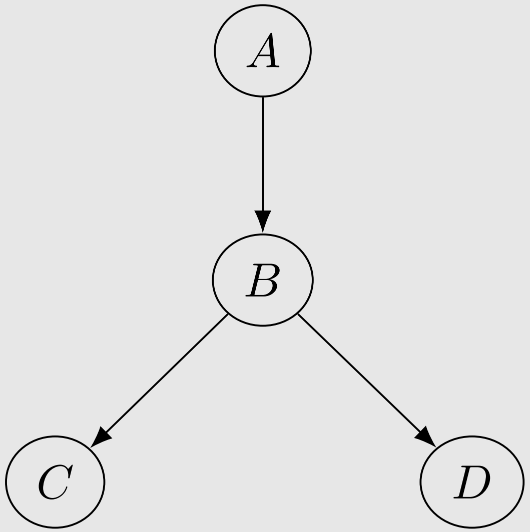
Question:
What graphs is the following graph Markov equivalent to?

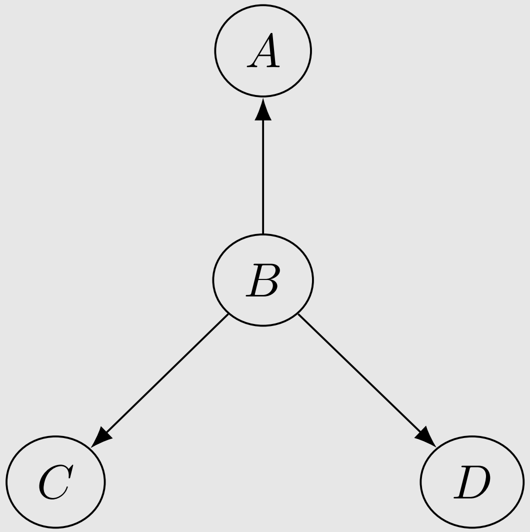


Question:
What graphs is the following graph Markov equivalent to?

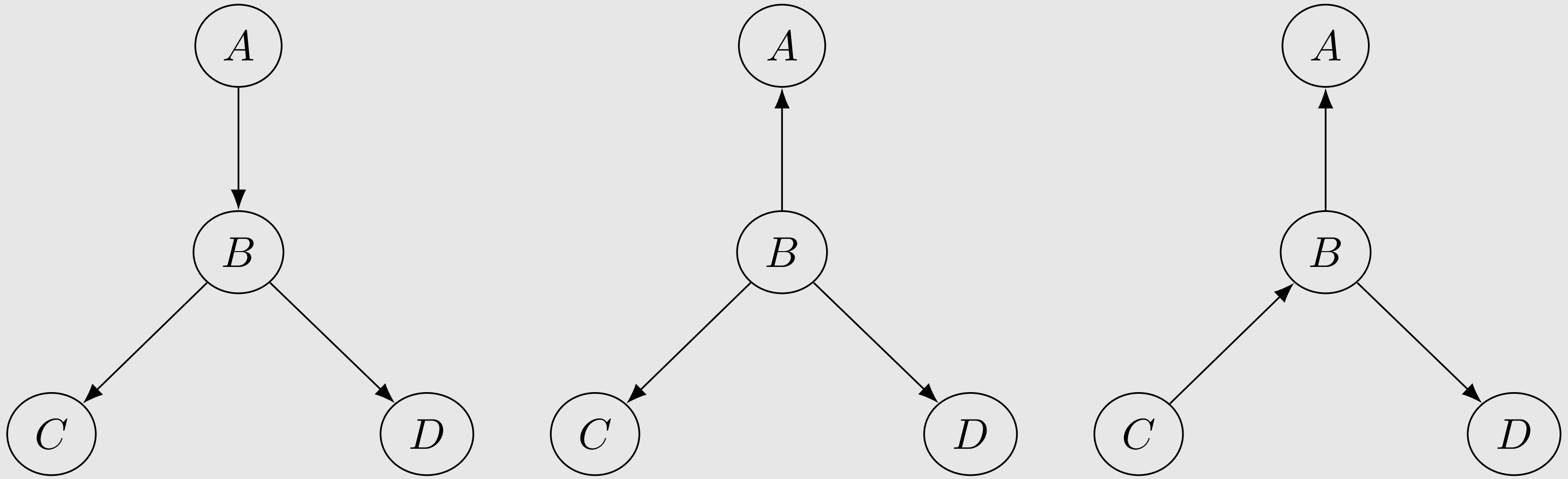


Question:
What graphs is the following graph Markov equivalent to?

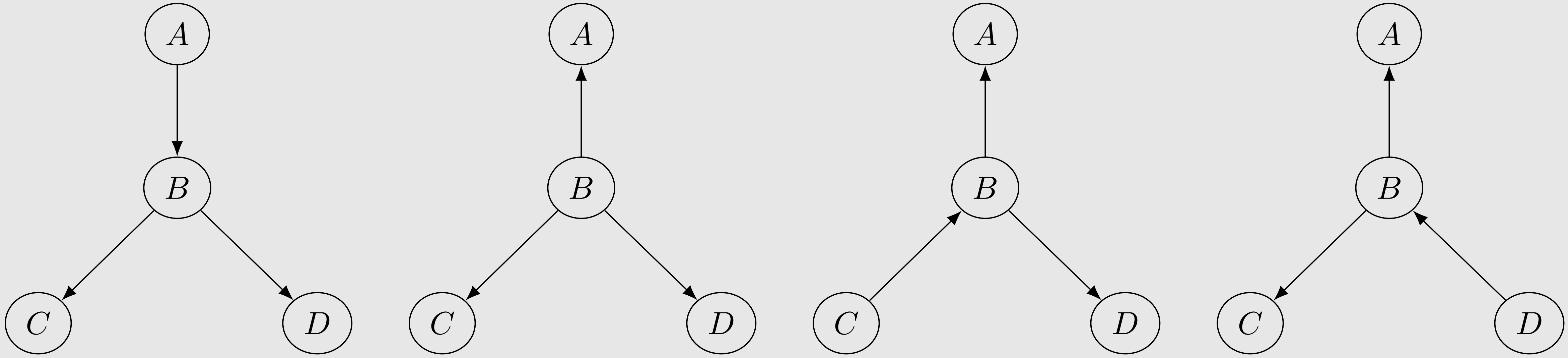




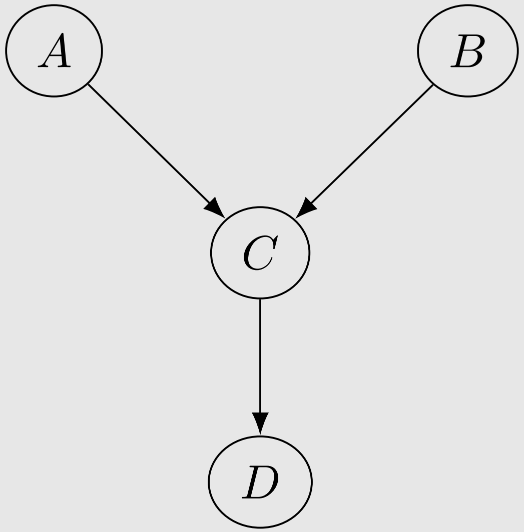
Question:
What graphs is the following graph Markov equivalent to?



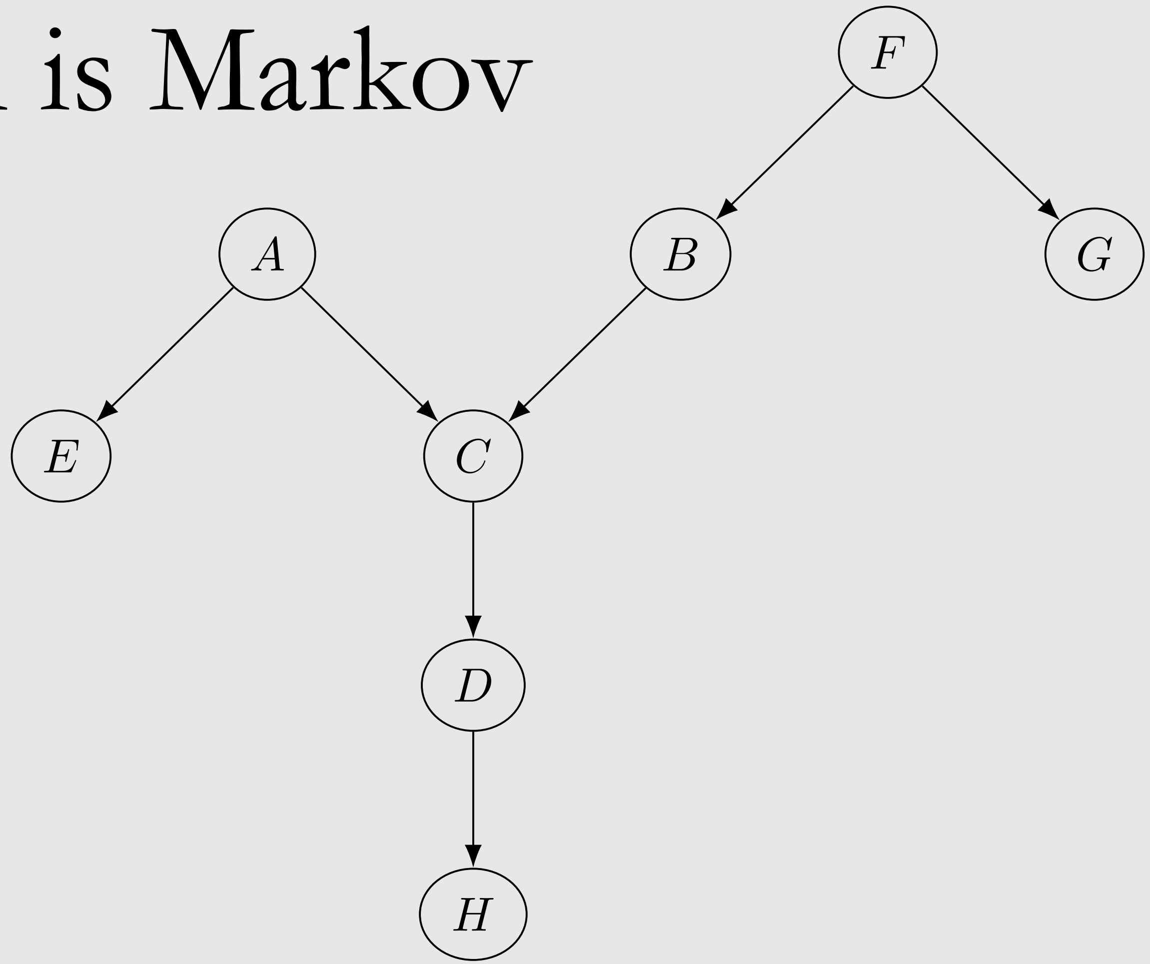
Question:
What graphs is the following graph Markov equivalent to?



Question:
What graphs is the following graph Markov equivalent to?



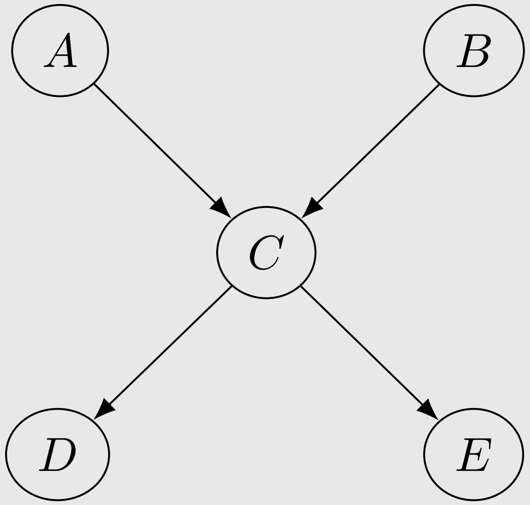
Question:
Give a few graphs that the following graph is Markov equivalent to:



The PC Algorithm: Overview

The PC Algorithm: Overview

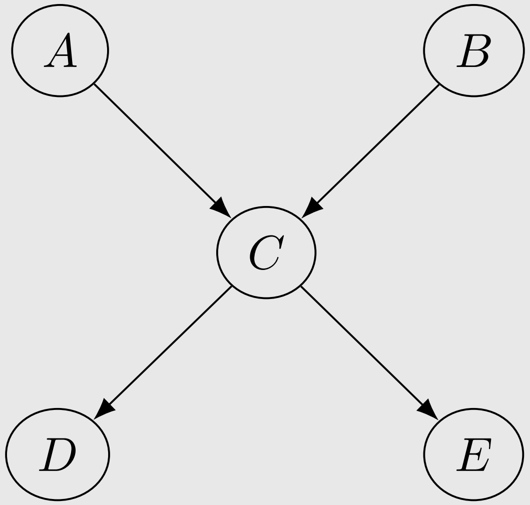
True graph



The PC Algorithm: Overview

Start with complete undirected graph

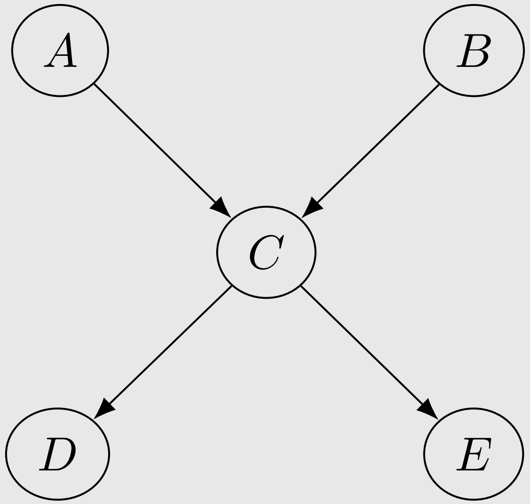
True graph

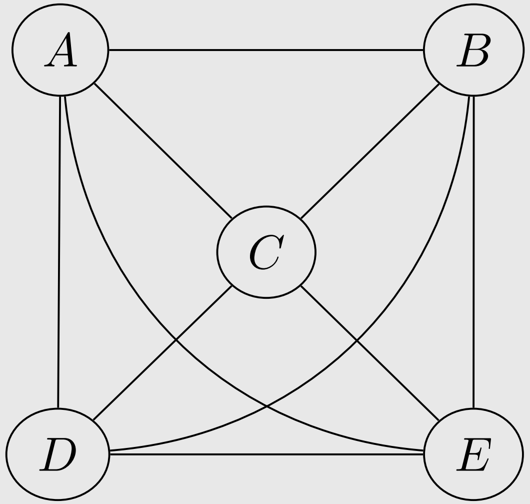


The PC Algorithm: Overview

Start with complete undirected graph

True graph



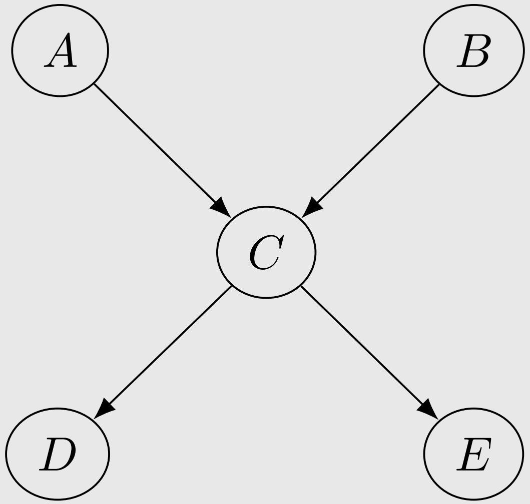


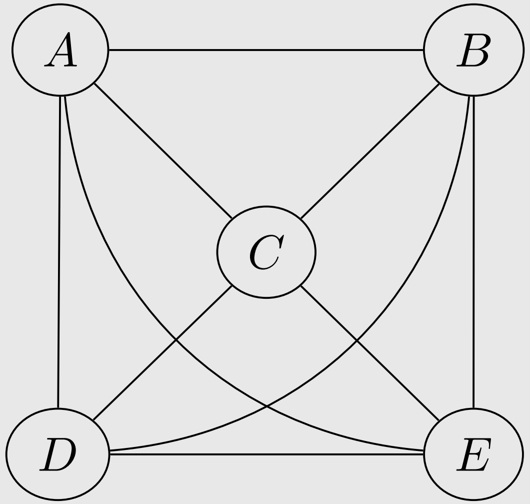
The PC Algorithm: Overview

Start with complete undirected graph

Three steps:

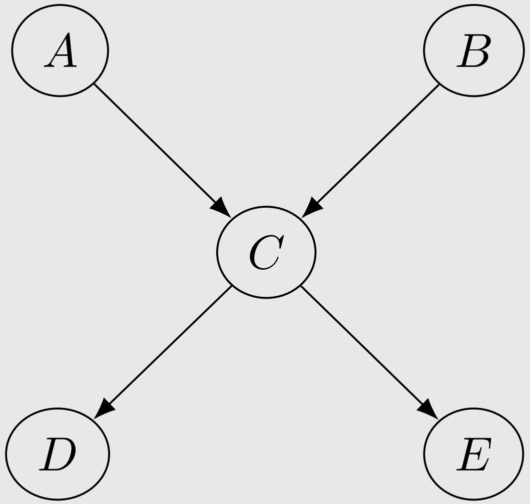
True graph





The PC Algorithm: Overview

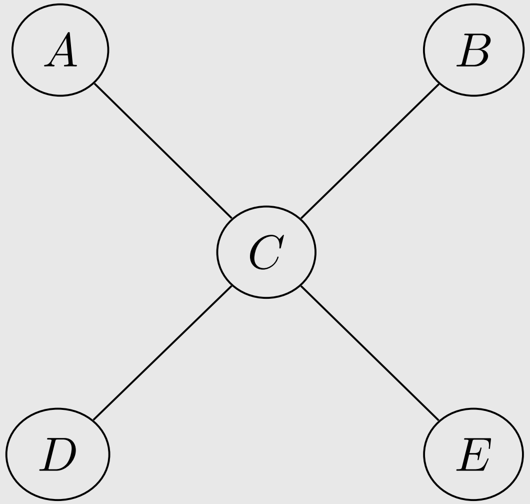
True graph



Start with complete undirected graph

Three steps:

1. Identify the skeleton



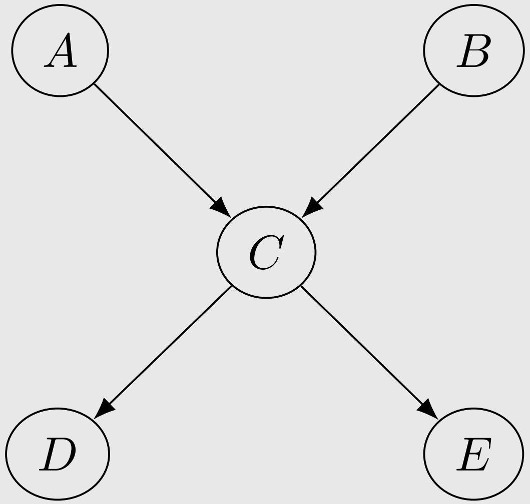
The PC Algorithm: Overview

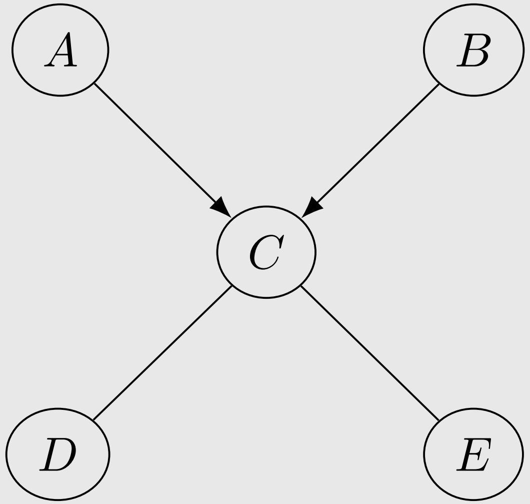
Start with complete undirected graph

Three steps:

1. Identify the skeleton
2. Identify immoralities and orient them

True graph





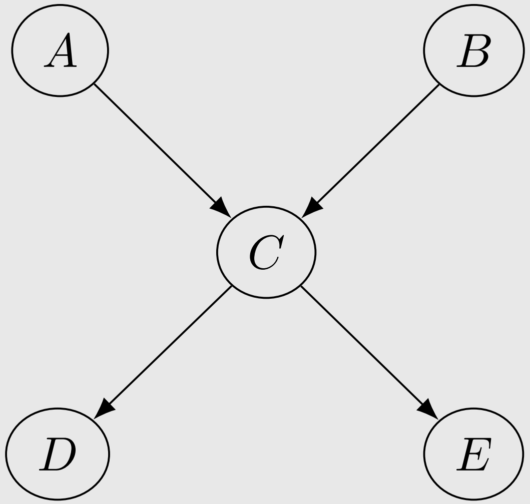
The PC Algorithm: Overview

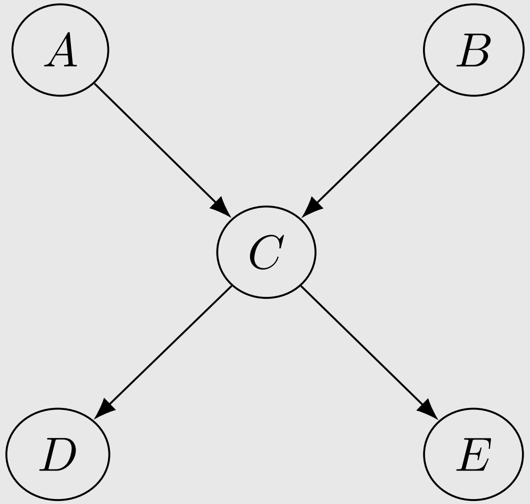
Start with complete undirected graph

Three steps:

1. Identify the skeleton
2. Identify immoralities and orient them
3. Orient qualifying edges that are incident on colliders

True graph





Identifying the Skeleton

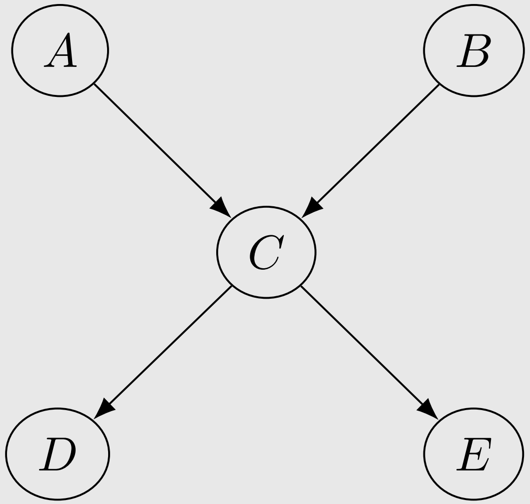
Identifying the Skeleton

Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

Identifying the Skeleton

Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

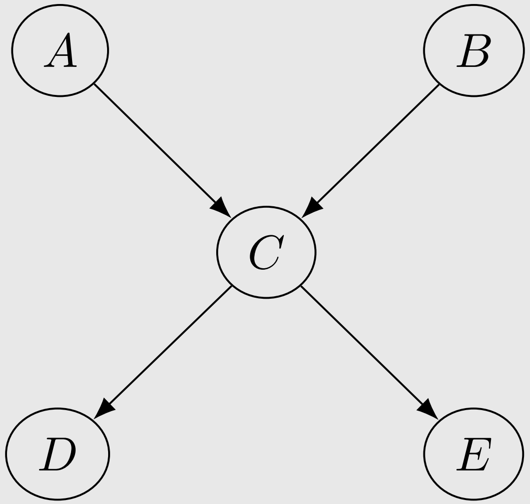
True graph

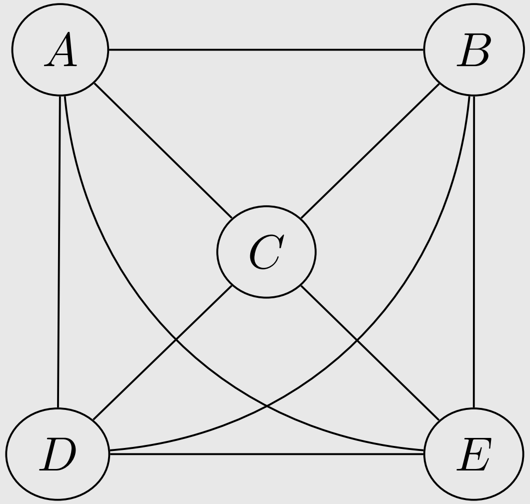


Identifying the Skeleton

Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

True graph

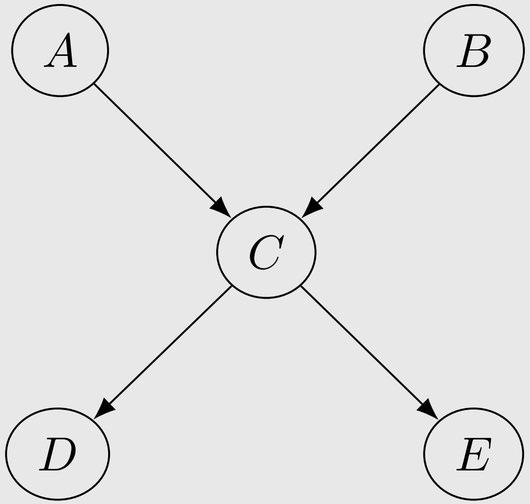




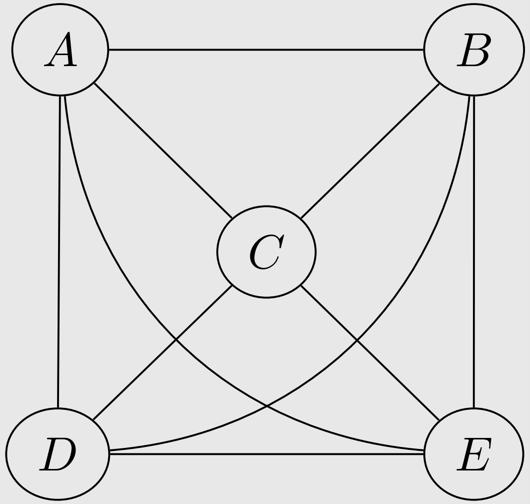
Identifying the Skeleton

Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

True graph



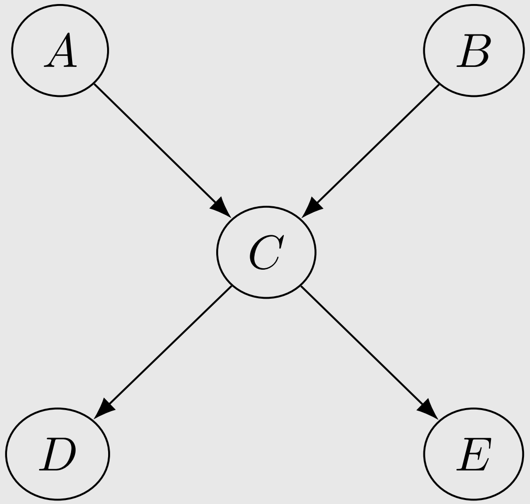
$$A \perp\!\!\!\perp B$$



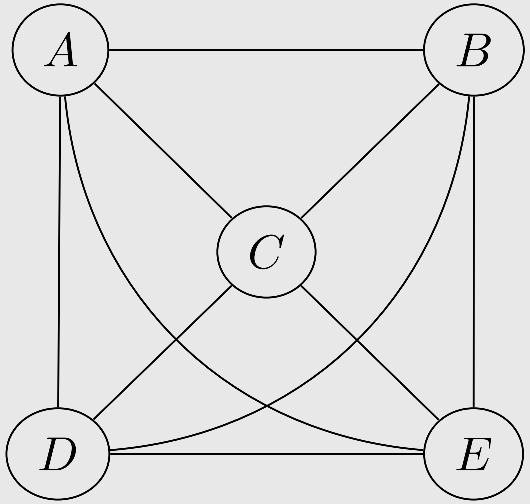
Identifying the Skeleton

Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

True graph



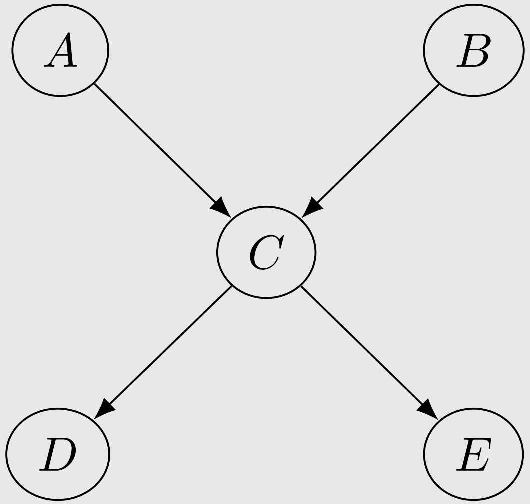
$$A \perp\!\!\!\perp B \mid \{\}$$



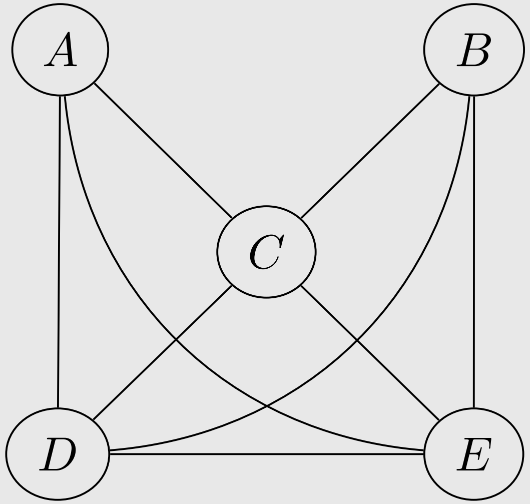
Identifying the Skeleton

Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

True graph



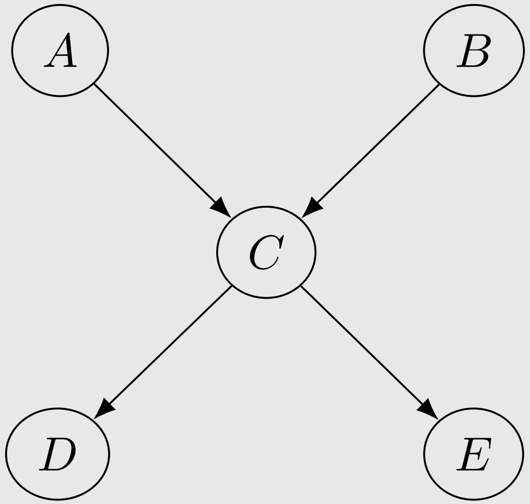
$$A \perp\!\!\!\perp B \mid \{\}$$



Identifying the Skeleton

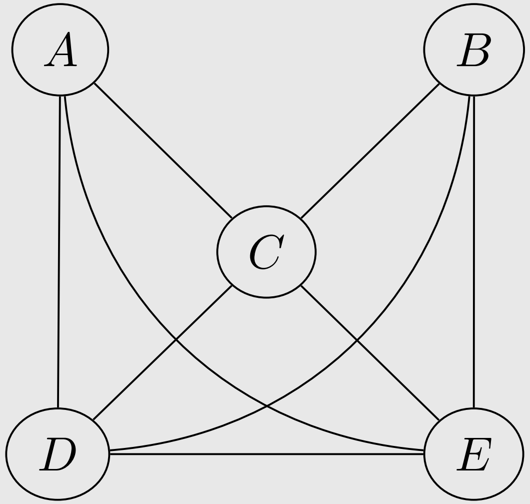
Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

True graph

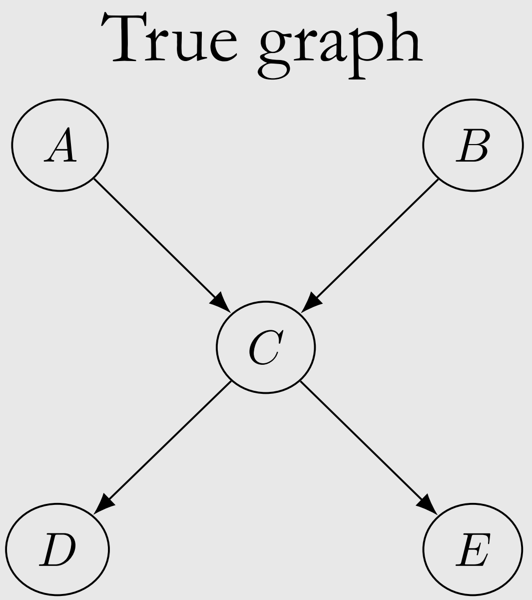


$$A \perp\!\!\!\perp B \mid \{\}$$

$$\forall \text{ other pairs } (X, Y), \quad X \perp\!\!\!\perp Y \mid \{C\}$$



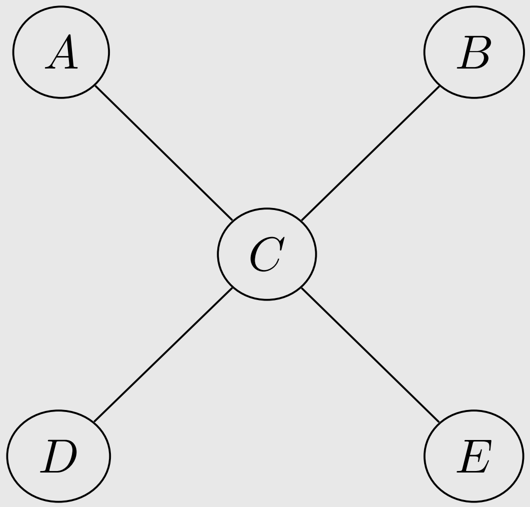
Identifying the Skeleton



Start with complete undirected graph and remove edges X – Y where $X \perp\!\!\!\perp Y \mid Z$ for some (potentially empty) conditioning set Z, starting with the empty conditioning set and increasing the size

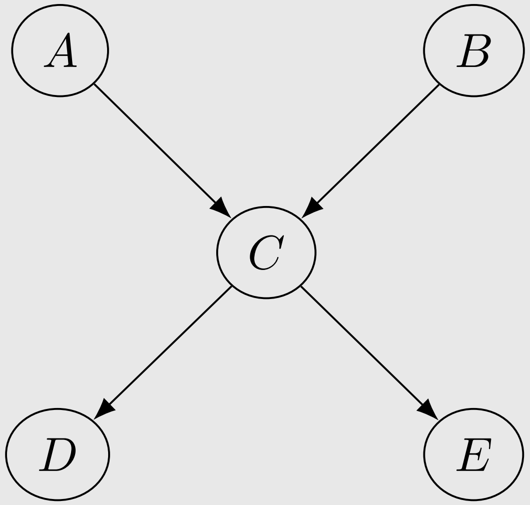
$$A \perp\!\!\!\perp B \mid \{\}$$

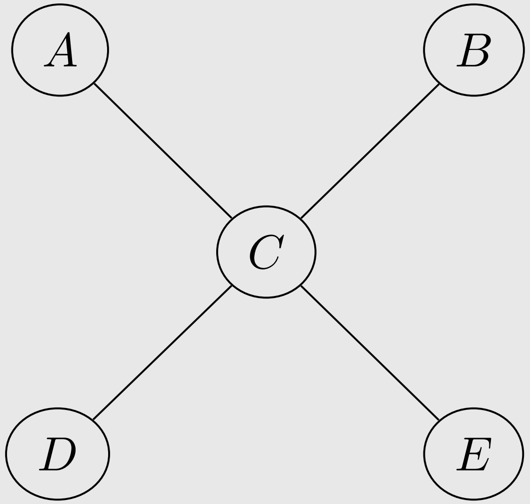
$$\forall \text{ other pairs } (X, Y), \quad X \perp\!\!\!\perp Y \mid \{C\}$$



Identifying the Immoralities

True graph

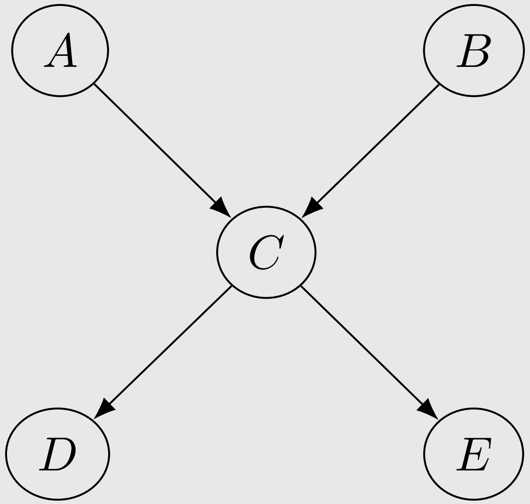


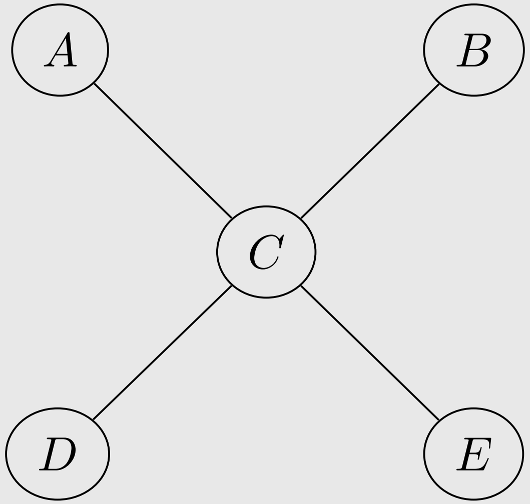


Identifying the Immoralities

Now for any paths X – Z – Y in our working graph where the following are true:

True graph



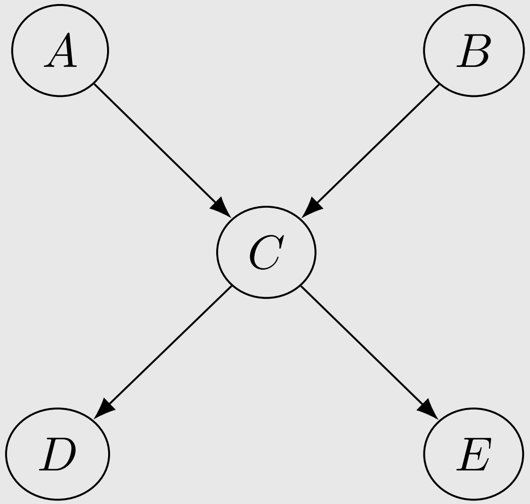


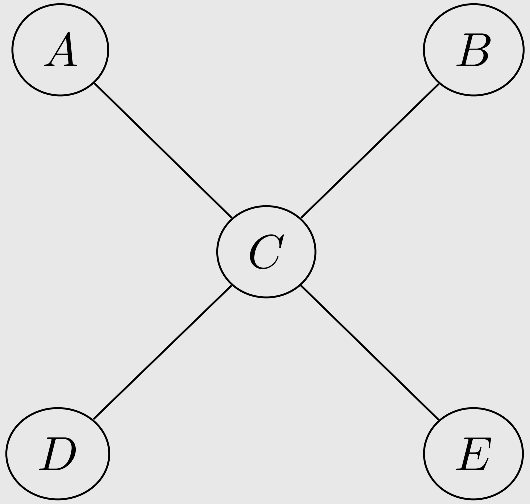
Identifying the Immoralities

Now for any paths X – Z – Y in our working graph where the following are true:

1. We discovered that there is no edge between X and Y in our previous step.

True graph



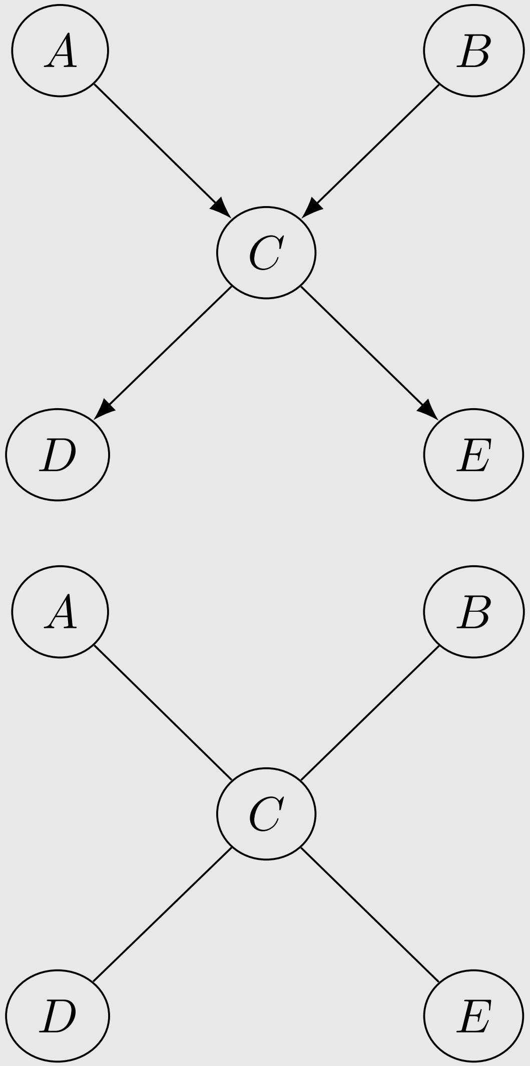


Identifying the Immoralities

Now for any paths X – Z – Y in our working graph where the following are true:

1. We discovered that there is no edge between X and Y in our previous step.
2. Z was not in the conditioning set that makes X and Y conditionally independent.

True graph



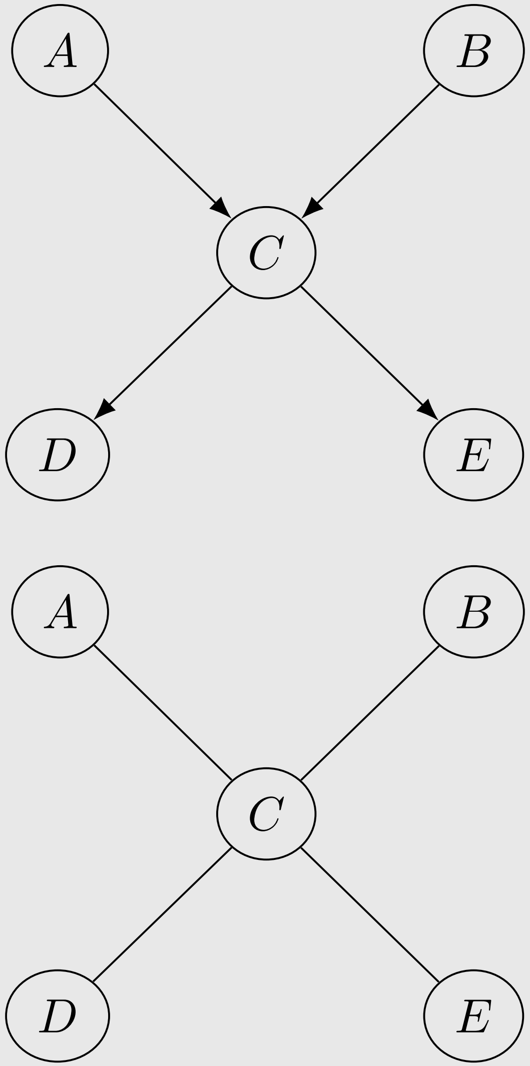
Identifying the Immoralities

Now for any paths X – Z – Y in our working graph where the following are true:

1. We discovered that there is no edge between X and Y in our previous step.
2. Z was not in the conditioning set that makes X and Y conditionally independent.

Then, we know X – Z – Y forms an immortality.

True graph



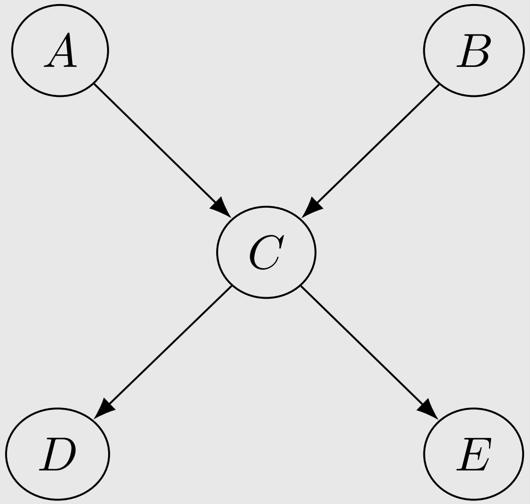
Identifying the Immoralities

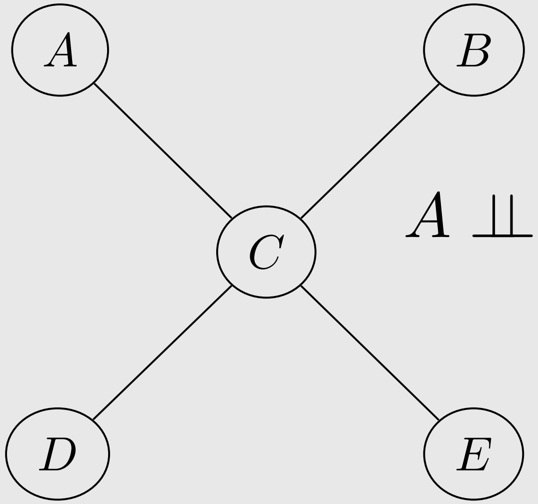
Now for any paths X – Z – Y in our working graph where the following are true:

1. We discovered that there is no edge between X and Y in our previous step.
2. Z was not in the conditioning set that makes X and Y conditionally independent.

Then, we know X – Z – Y forms an immortality.

True graph





$A \perp\!\!\!\perp B \mid \{\}$

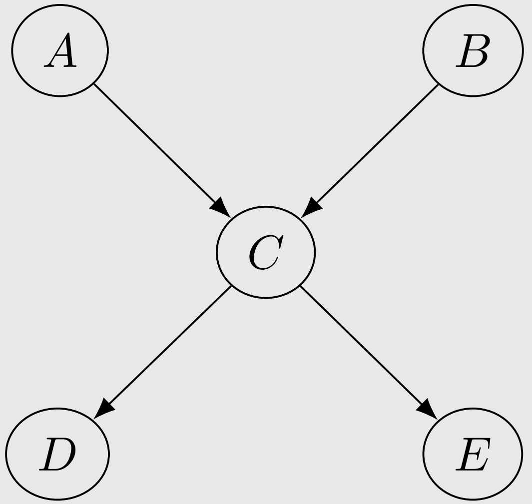
Identifying the Immoralities

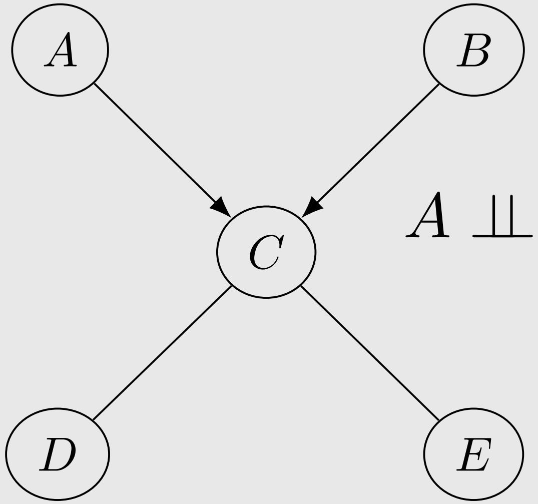
Now for any paths X – Z – Y in our working graph where the following are true:

1. We discovered that there is no edge between X and Y in our previous step.
2. Z was not in the conditioning set that makes X and Y conditionally independent.

Then, we know X – Z – Y forms an immortality.

True graph

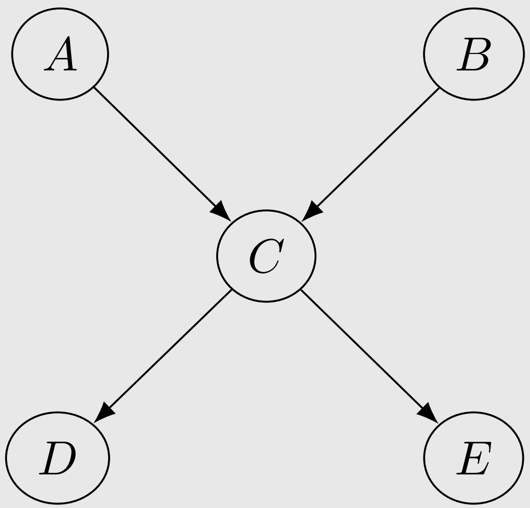


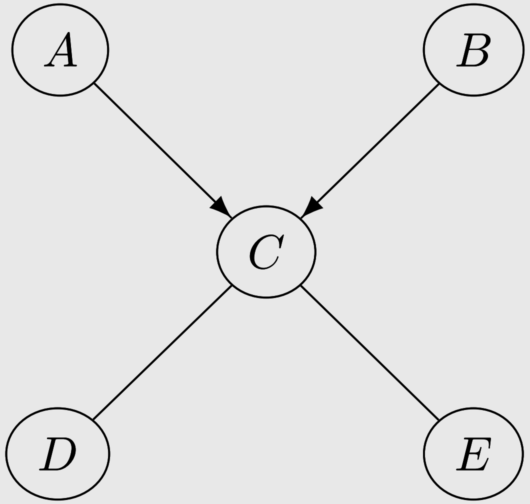


$A \perp\!\!\!\perp B \mid \{\}$

Orienting Qualifying Edges Incident on Colliders

True graph

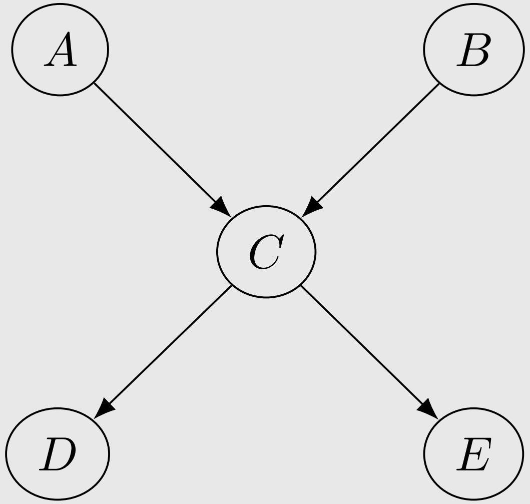


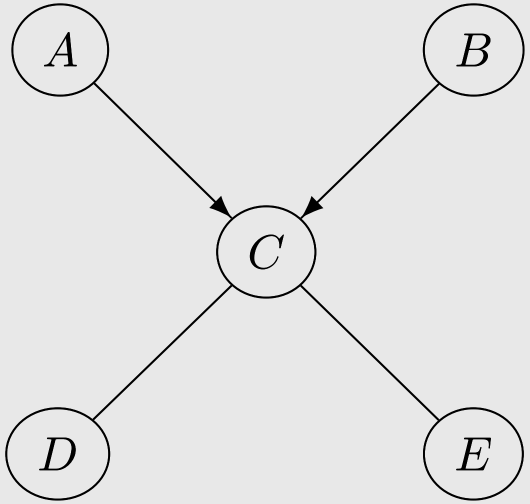


Orienting Qualifying Edges Incident on Colliders

Idea: use fact that we discovered all immoralities to orient more edges

True graph



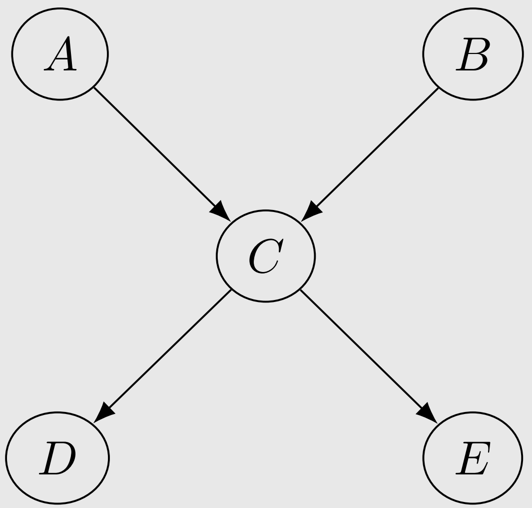


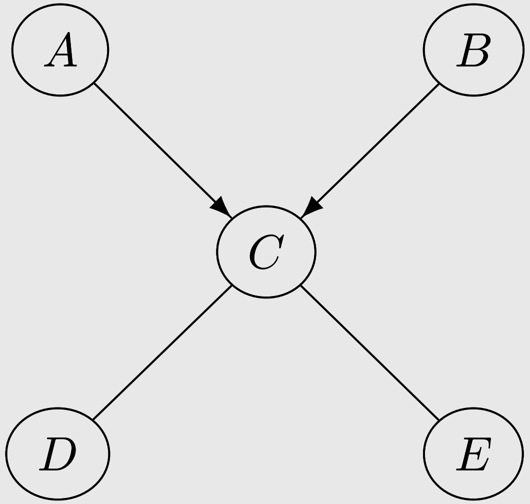
Orienting Qualifying Edges Incident on Colliders

Idea: use fact that we discovered all immoralities to orient more edges

Any edge $Z - Y$ part of a partially directed path of the form $X \to Z - Y$, where there is no edge connecting X and Y can be oriented as $Z \to Y$

True graph



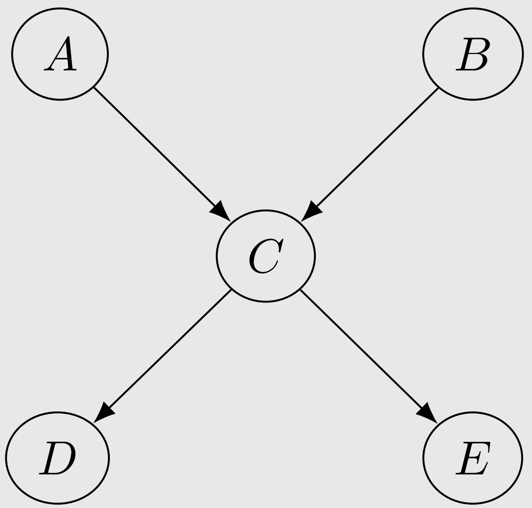


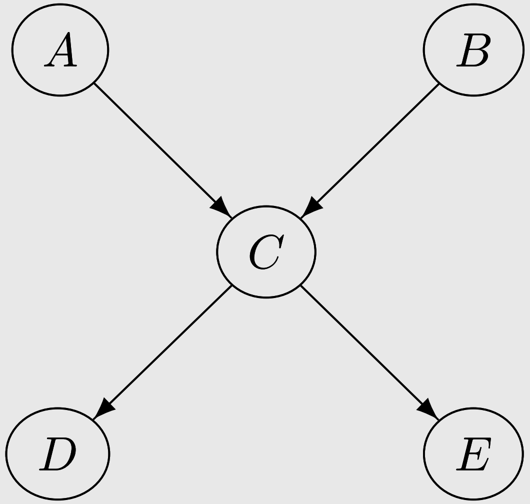
Orienting Qualifying Edges Incident on Colliders

Idea: use fact that we discovered all immoralities to orient more edges

Any edge $Z - Y$ part of a partially directed path of the form $X \rightarrow Z - Y$, where there is no edge connecting X and Y can be oriented as $Z \rightarrow Y$

True graph





Removing Assumptions

Removing Assumptions

No assumed causal sufficiency: FCI algorithm (Spirtes et al., 2001)

Removing Assumptions

No assumed causal sufficiency: FCI algorithm (Spirtes et al., 2001)

No assumed acyclicity: CCD algorithm (Richardson, 1996)

Removing Assumptions

No assumed causal sufficiency: FCI algorithm (Spirtes et al., 2001)

No assumed acyclicity: CCD algorithm (Richardson, 1996)

Neither causal sufficiency nor acyclicity: SAT-based causal discovery (Hyttinen et al., 2013; 2014)

Hardness of Conditional Independence Testing

Hardness of Conditional Independence Testing

Independence-based causal discovery algorithms rely on accurate conditional independence testing.

Hardness of Conditional Independence Testing

Independence-based causal discovery algorithms rely on accurate conditional independence testing.

Conditional independence testing is simple if we have infinite data.

Hardness of Conditional Independence Testing

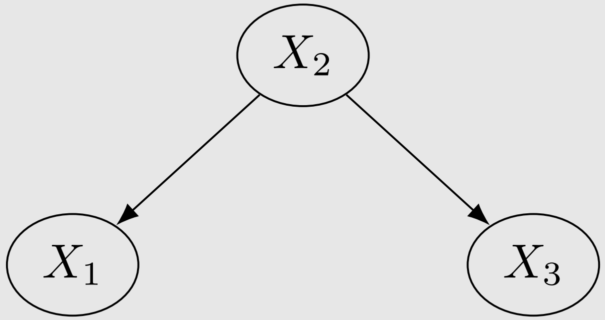
Independence-based causal discovery algorithms rely on accurate conditional independence testing.

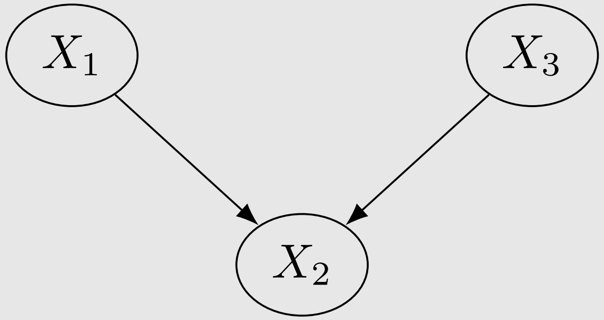
Conditional independence testing is simple if we have infinite data.

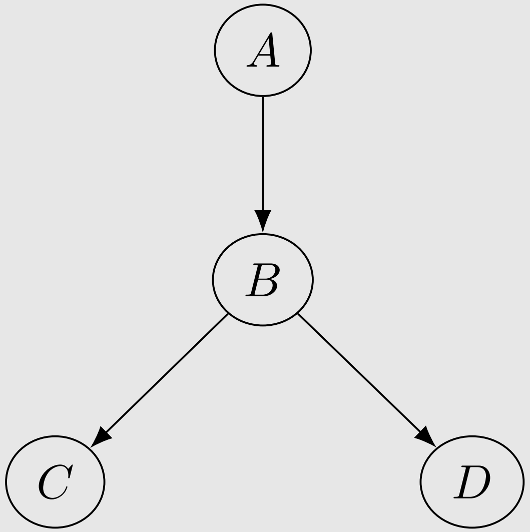
However, it is a quite hard problem with finite data, and it can sometimes require a lot of data to get accurate test results (Shah & Peters, 2020).

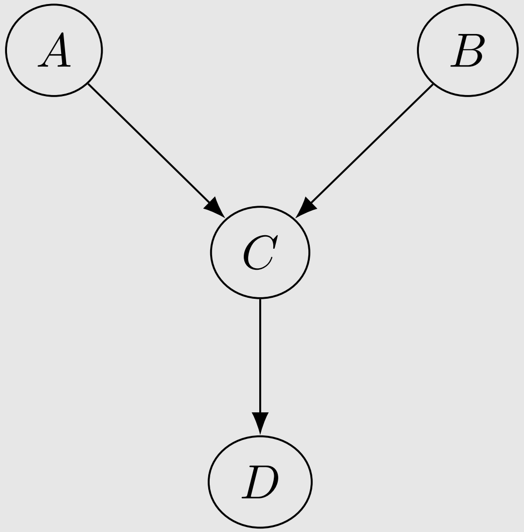
Questions:

1. What are the essential graphs of the following graphs?



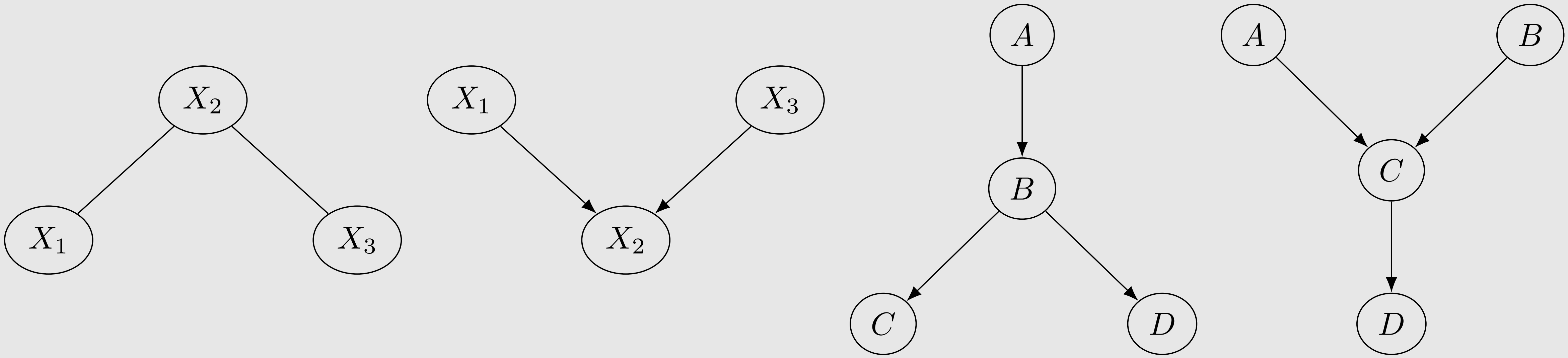






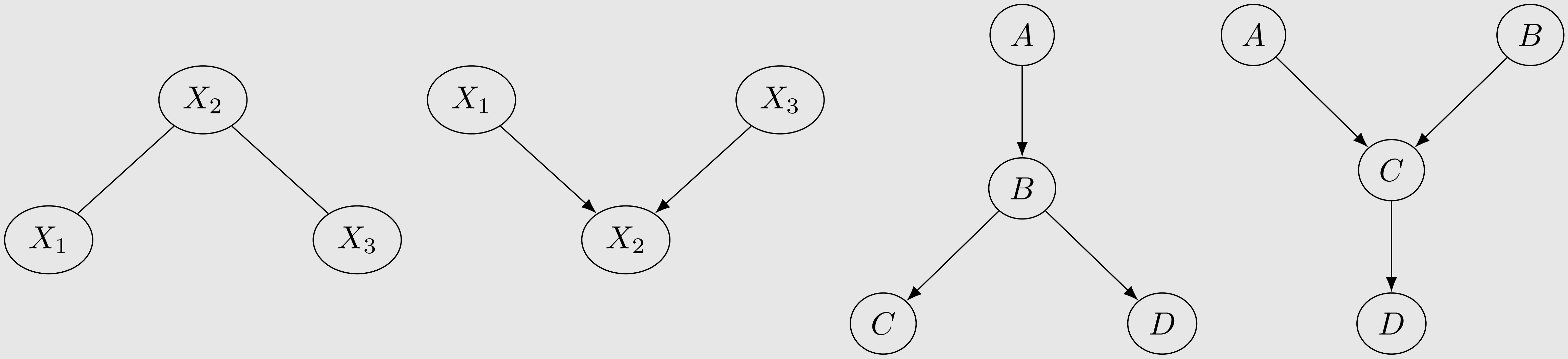
Questions:

1. What are the essential graphs of the following graphs?



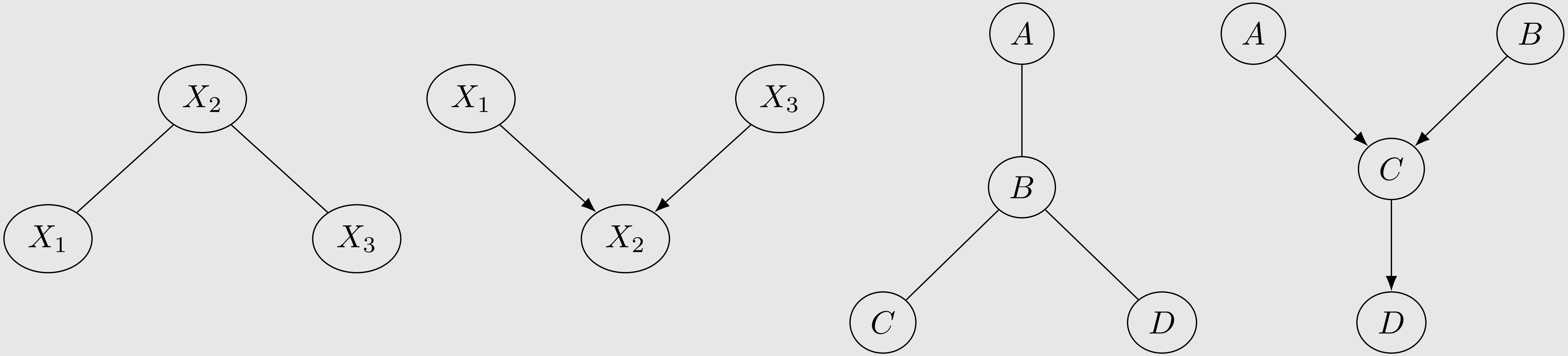
Questions:

1. What are the essential graphs of the following graphs?



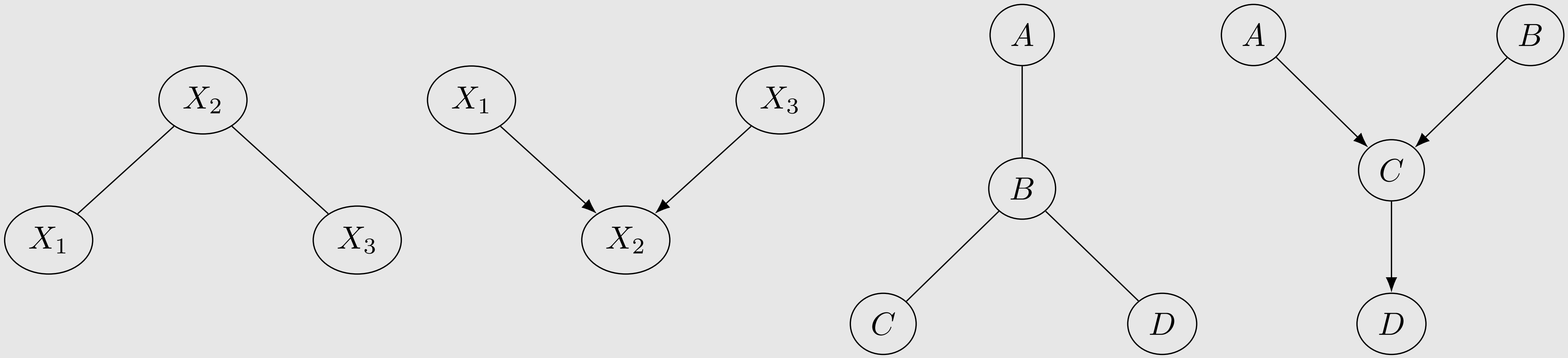
Questions:

1. What are the essential graphs of the following graphs?



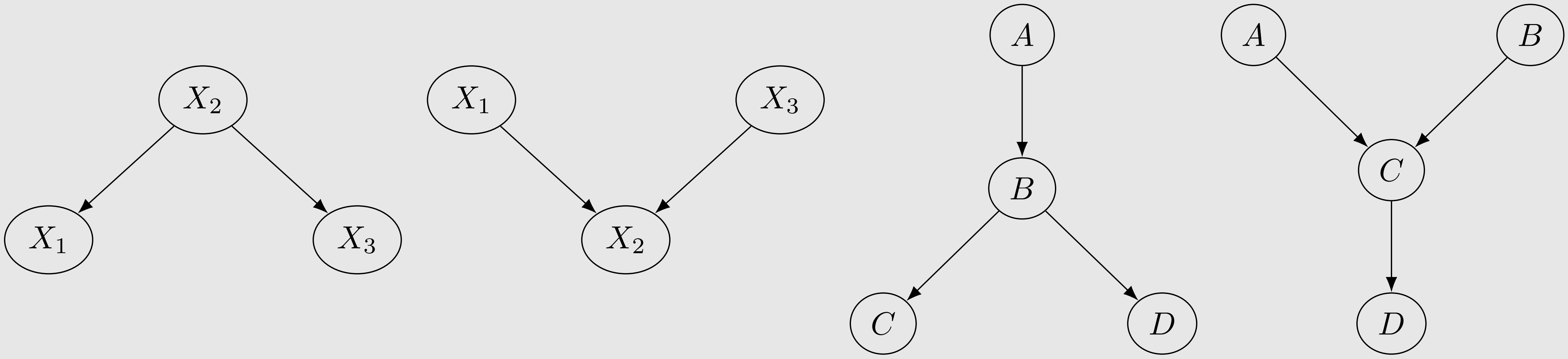
Questions:

1. What are the essential graphs of the following graphs?

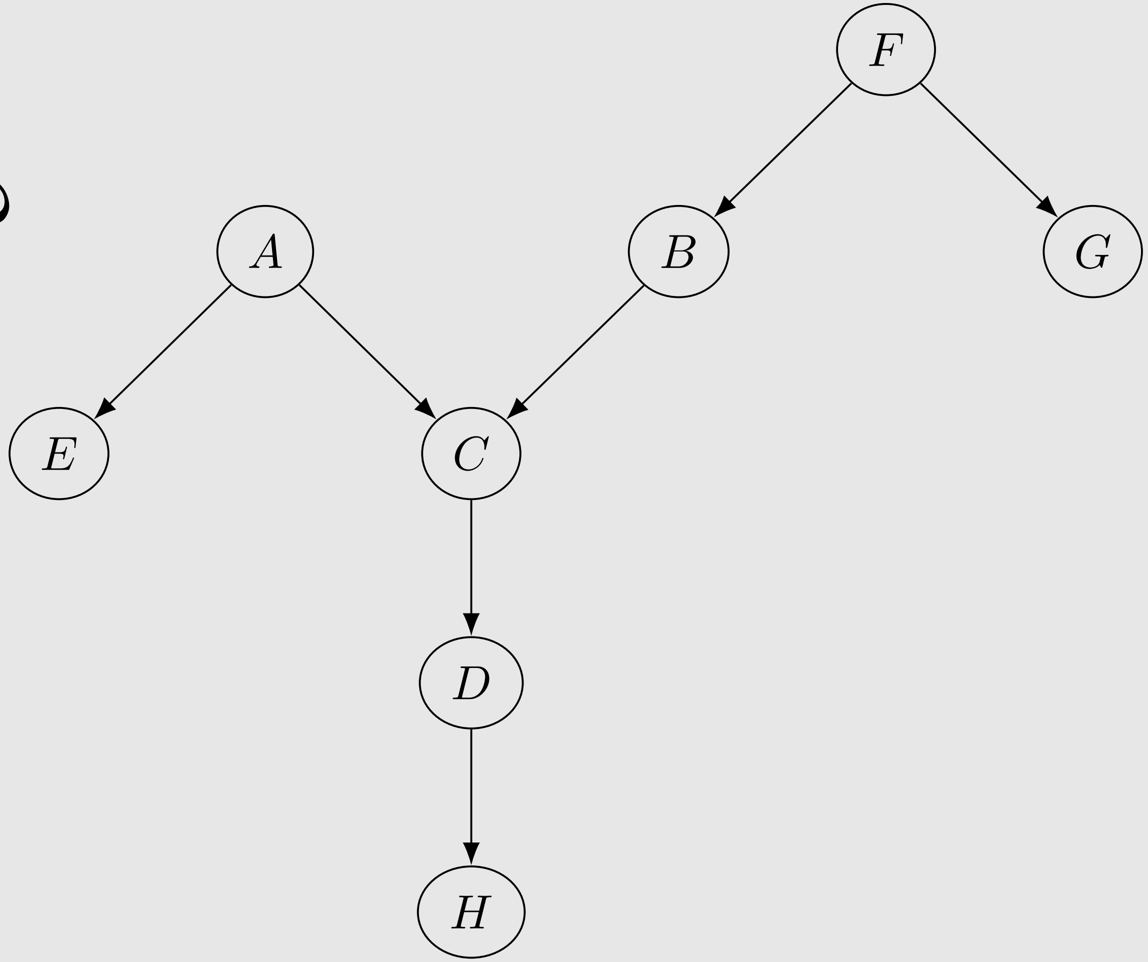


Questions:

1. What are the essential graphs of the following graphs?
2. Walk through the steps of PC to get them.



Question:
What is the essential graph for this graph?



Can We Do Better?

With faithfulness, we saw we can identify the essential graph (Markov equivalence class).

Can We Do Better?

With faithfulness, we saw we can identify the essential graph (Markov equivalence class).

If we have multinomial distributions (Meek, 1995) or linear Gaussian structural equations (Geiger & Pearl, 1988), we can only identify a graph up to its Markov equivalence class.

Can We Do Better?

With faithfulness, we saw we can identify the essential graph (Markov equivalence class).

If we have multinomial distributions (Meek, 1995) or linear Gaussian structural equations (Geiger & Pearl, 1988), we can only identify a graph up to its Markov equivalence class.

What about non-Gaussian structural equations?

Or nonlinear structural equations?

Independence-Based Causal Discovery

Semi-Parametric Causal Discovery

Issues with Independence-Based Causal Discovery

Issues with Independence-Based Causal Discovery

- Requires faithfulness assumption

Issues with Independence-Based Causal Discovery

- Requires faithfulness assumption
- Large samples can be necessary for conditional independence tests

Issues with Independence-Based Causal Discovery

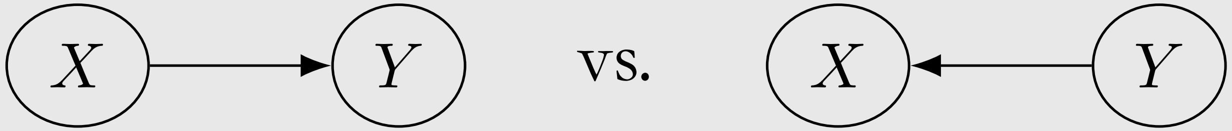
- Requires faithfulness assumption
- Large samples can be necessary for conditional independence tests
- Only identifies the Markov equivalence class

Two Variable Case: Markov Equivalence

Infinite data: $P(x, y)$

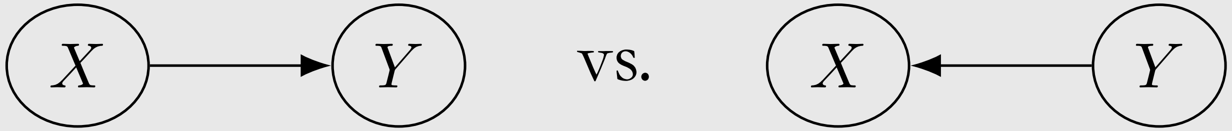
Two Variable Case: Markov Equivalence

Infinite data: $P(x, y)$

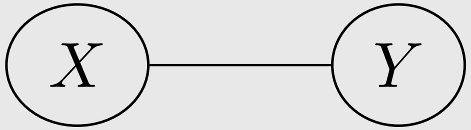


Two Variable Case: Markov Equivalence

Infinite data: $P(x, y)$



Essential graph:



Two Variable Case: SCMs Perspective

Two Variable Case: SCMs Perspective

Proposition: For every joint distribution $P(x, y)$ on two real-valued random variables, there is an SCM in either direction that generates data consistent with $P(x, y)$.

Two Variable Case: SCMs Perspective

Proposition: For every joint distribution $P(x, y)$ on two real-valued random variables, there is an SCM in either direction that generates data consistent with $P(x, y)$.

Mathematically, there exists a function f_Y such that

$$Y = f_Y(X, U_Y), \quad X \perp\!\!\!\perp U_Y$$

and there exists a function f_X such that

$$X = f_X(Y, U_X), \quad Y \perp\!\!\!\perp U_X$$

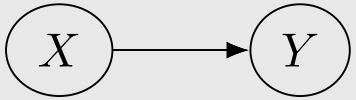
where U_Y and U_X are real-valued random variables.

Two Variable Case: SCMs Perspective

Proposition: For every joint distribution $P(x, y)$ on two real-valued random variables, there is an SCM in either direction that generates data consistent with $P(x, y)$.

Mathematically, there exists a function f_Y such that

$$Y = f_Y(X, U_Y)\,, \quad X \perp\!\!\!\perp U_Y$$



and there exists a function f_X such that

$$X = f_X(Y, U_X)\,, \quad Y \perp\!\!\!\perp U_X$$

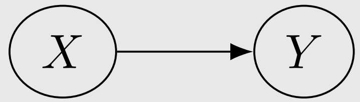
where U_Y and U_X are real-valued random variables.

Two Variable Case: SCMs Perspective

Proposition: For every joint distribution $P(x, y)$ on two real-valued random variables, there is an SCM in either direction that generates data consistent with $P(x, y)$.

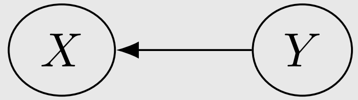
Mathematically, there exists a function f_Y such that

$$Y = f_Y(X, U_Y)\,, \quad X \perp\!\!\!\perp U_Y$$



and there exists a function f_X such that

$$X = f_X(Y, U_X)\,, \quad Y \perp\!\!\!\perp U_X$$



where U_Y and U_X are real-valued random variables.

We must make assumptions about the parametric form.

Linear Non-Gaussian Assumption

Linear Non-Gaussian Assumption

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

Linear Non-Gaussian Assumption

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

What if the noise is non-Gaussian?

Linear Non-Gaussian Assumption

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

What if the noise is non-Gaussian?

Linear Non-Gaussian Assumption:

All structural equations (causal mechanisms that generate the data) are of the following form:

$$Y := f(X) + U$$

where f is a linear function, $X \perp\!\!\!\perp U$, and U is distributed as some non-Gaussian.

Identifiability in Linear Non-Gaussian Setting

Theorem (Shimizu et al., 2006):

In the linear non-Gaussian setting, if the true SCM is

$$Y := f(X) + U\,, \quad X \perp\!\!\!\perp U\,,$$

then there does not exist an SCM in the reverse direction,

$$X := g(Y) + \tilde{U}\,, \quad Y \perp\!\!\!\perp \tilde{U}\,,$$

that can generate data consistent with $P(x, y)$.

Identifiability in Linear Non-Gaussian Setting

Theorem (Shimizu et al., 2006):

In the linear non-Gaussian setting, if the true SCM is

$$Y := f(X) + U \,, \quad X \perp\!\!\!\perp U \,,$$

then there does not exist an SCM in the reverse direction,

$$X := g(Y) + \tilde{U} \,, \quad Y \perp\!\!\!\perp \tilde{U} \,,$$

that can generate data consistent with $P(x, y)$.

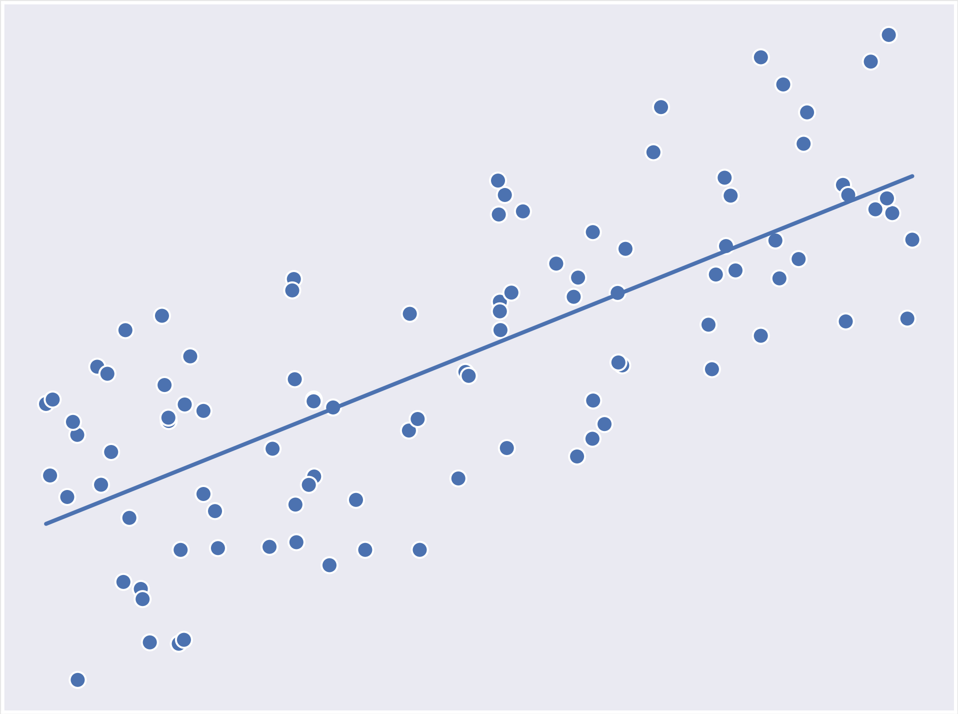
See proof in the course book

Identifiability in Linear Non-Gaussian Setting: Linear Fit

Identifiability in Linear Non-Gaussian Setting: Linear Fit

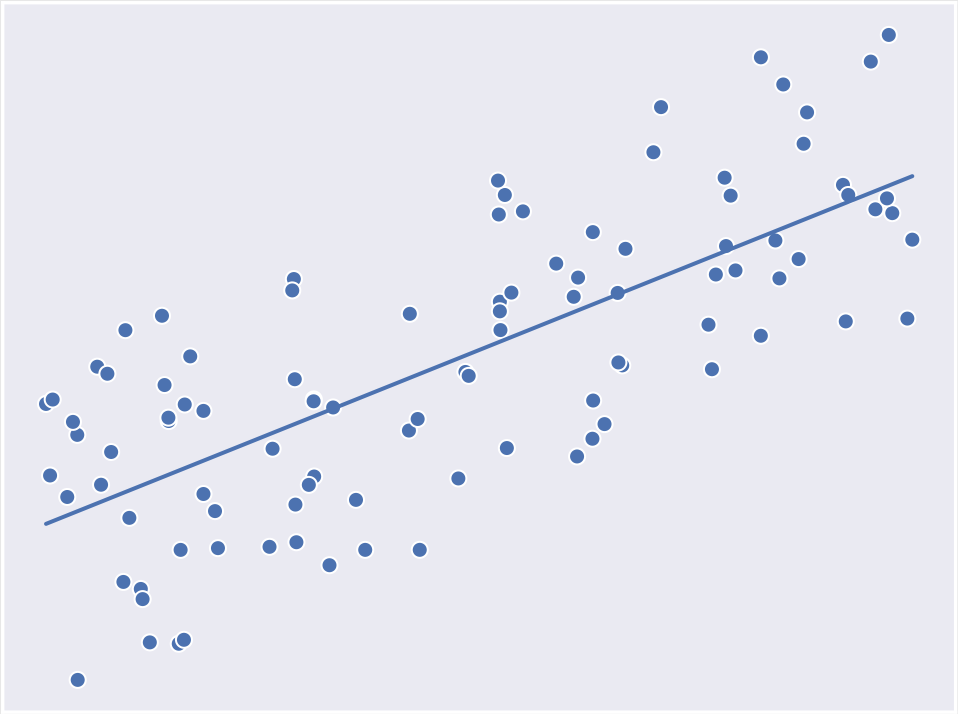
$$Y := f(X) + U$$

Identifiability in Linear Non-Gaussian Setting: Linear Fit



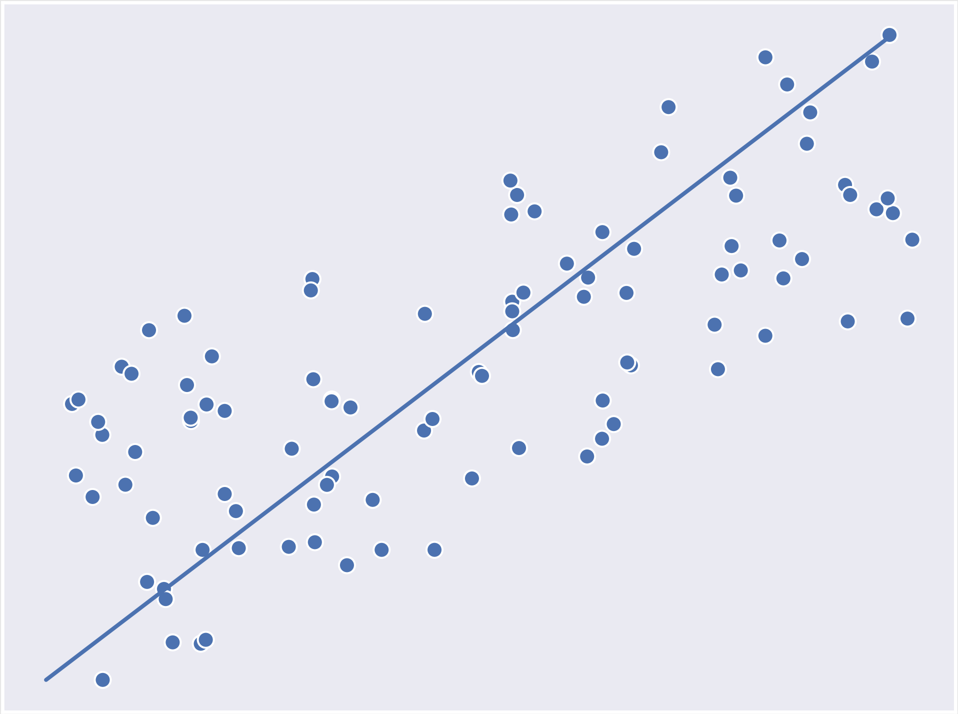
$$Y := f(X) + U$$

Identifiability in Linear Non-Gaussian Setting: Linear Fit



$$X := g(Y) + \tilde{U}$$

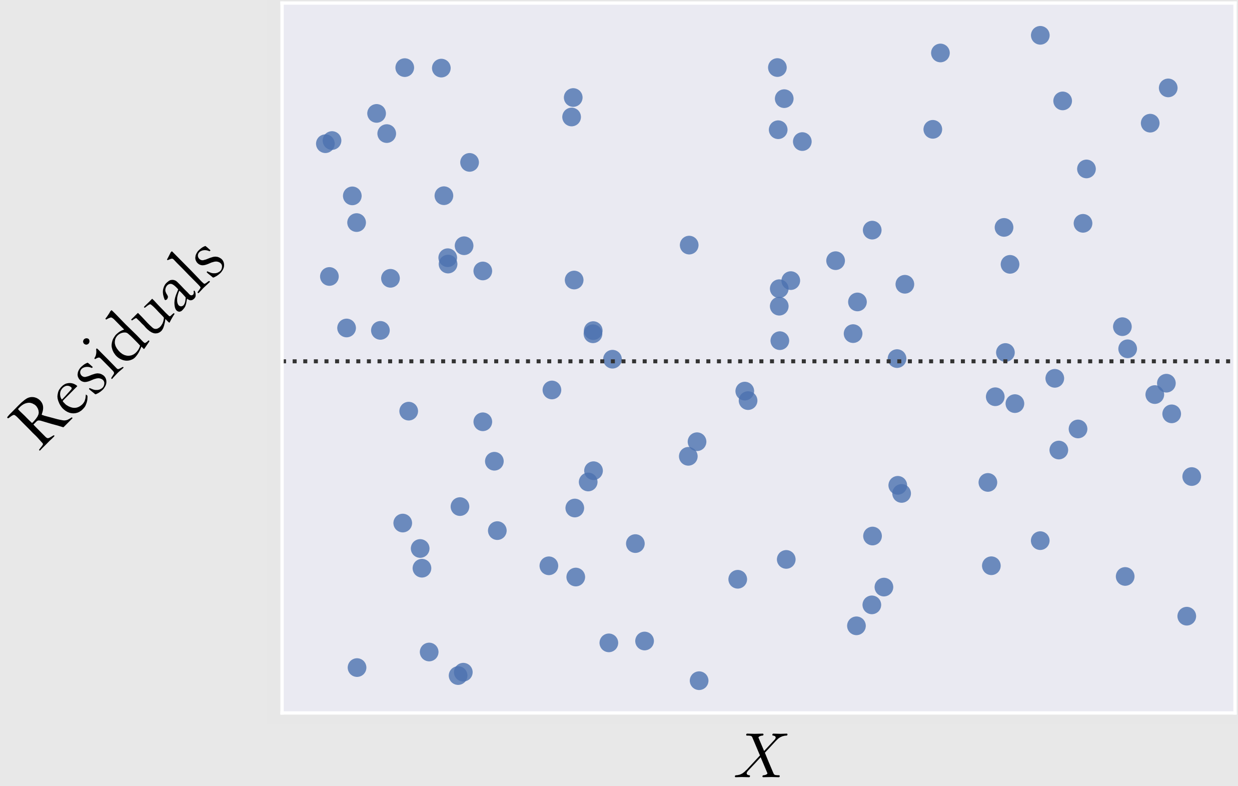
Identifiability in Linear Non-Gaussian Setting: Linear Fit



$$X := g(Y) + \tilde{U}$$

Identifiability in Linear Non-Gaussian Setting: Residuals

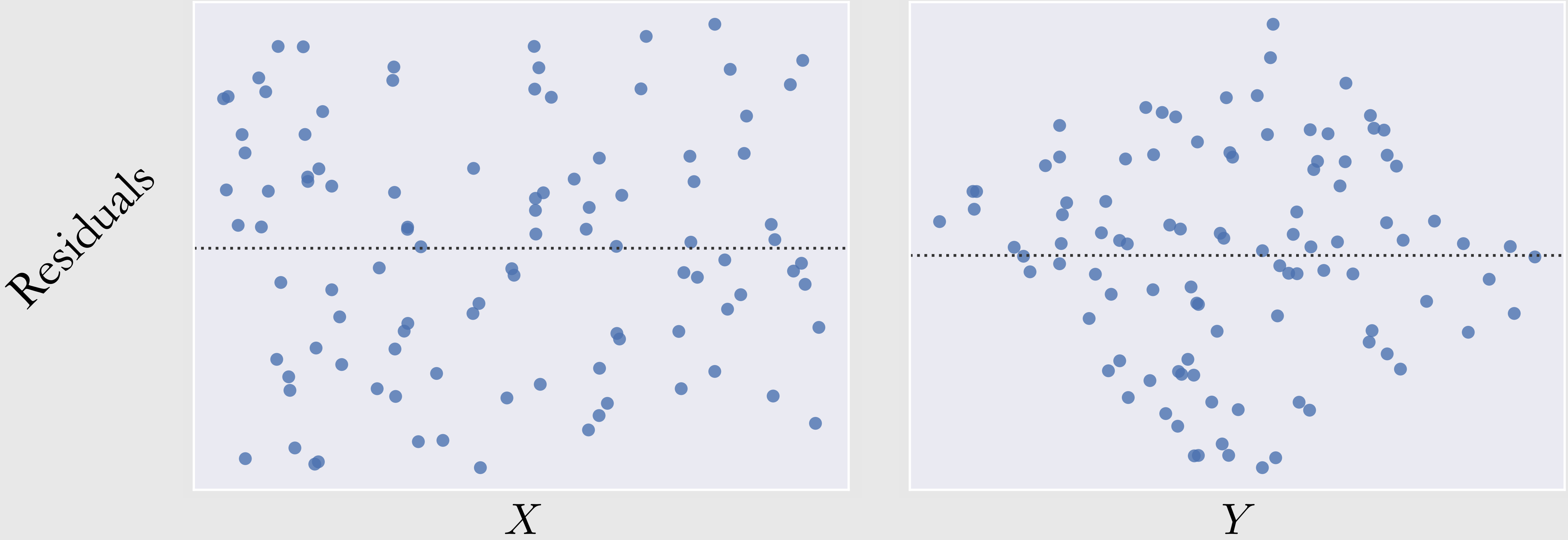
$$Y := f(X) + U\,, \quad X \perp\!\!\!\perp U$$



Identifiability in Linear Non-Gaussian Setting: Residuals

$$Y := f(X) + U \,, \quad X \perp\!\!\!\perp U$$

$$X := g(Y) + \tilde{U} \,, \quad Y \not\perp\!\!\!\perp \tilde{U}$$



Linear Non-Gaussian Identifiability Extensions

Linear Non-Gaussian Identifiability Extensions

- Multivariate: Shimizu et al., (2006)

Linear Non-Gaussian Identifiability Extensions

- Multivariate: Shimizu et al., (2006)

- Drop causal sufficiency assumption: Hoyer et al. (2008)

Linear Non-Gaussian Identifiability Extensions

- Multivariate: Shimizu et al., (2006)
- Drop causal sufficiency assumption: Hoyer et al. (2008)
- Drop acyclicity assumption: Lacerda et al. (2008)

Identifiability in Nonlinear Additive Noise Setting

Identifiability in Nonlinear Additive Noise Setting

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

Identifiability in Nonlinear Additive Noise Setting

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

What if the structural equations are nonlinear?

Identifiability in Nonlinear Additive Noise Setting

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

What if the structural equations are nonlinear?

Nonlinear additive noise assumption: $\forall i \,, \quad X_i := f_i(\text{pa}_i) + U_i$ where f_i is nonlinear

Identifiability in Nonlinear Additive Noise Setting

Recall: We cannot hope to identify the graph more precisely than the Markov equivalence class in the linear Gaussian noise setting (Geiger & Pearl, 1988).

What if the structural equations are nonlinear?

Nonlinear additive noise assumption: $\forall i\,, \quad X_i := f_i(\text{pa}_i) + U_i$ where f_i is nonlinear

Theorem (Hoyer et al. 2008): Under the Markov assumption, causal sufficiency, acyclicity, the nonlinear additive noise assumption, and a technical condition from Hoyer et al. (2008), we can identify the causal graph.

Post-Nonlinear Setting

Nonlinear additive noise setting: $Y := f(X) + U, \quad X \perp\!\!\!\perp U$

Post-Nonlinear Setting

Nonlinear additive noise setting: $Y := f(X) + U\,, \quad X \perp\!\!\!\perp U$

$$Y := f(X) + U\,, \quad X \perp\!\!\!\perp U$$

Post-Nonlinear Setting

Nonlinear additive noise setting: $Y := f(X) + U \,, \quad X \perp\!\!\!\perp U$

$$Y := g(f(X) + U)\,, \quad X \perp\!\!\!\perp U$$

Post-Nonlinear Setting

Nonlinear additive noise setting: $Y := f(X) + U \,, \quad X \perp\!\!\!\perp U$

Post-nonlinear (Zhang & Hyvärinen, 2009):

$$Y := g(f(X) + U) \,, \quad X \perp\!\!\!\perp U$$

Review Articles and Book

- Introduction to the Foundations of Causal Discovery (Eberhardt, 2017)
- Review of Causal Discovery Methods Based on Graphical Models (Glymour, Zhang, & Spirtes, 2019)
- Elements of Causal Inference (Peters, Janzing, & Schölkopf, 2017)