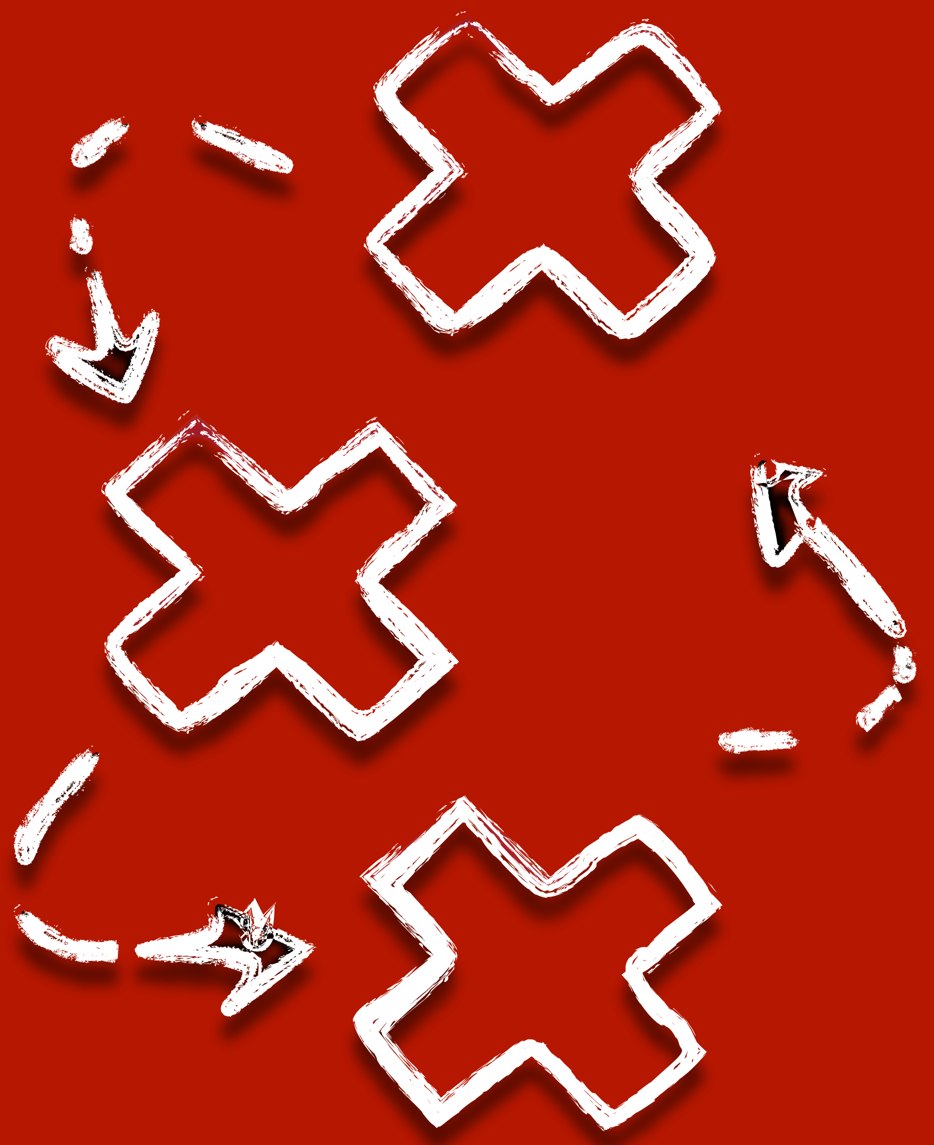


Causal Data Science

Lecture 3.2: Bayesian networks

Lecturer: Sara Magliacane

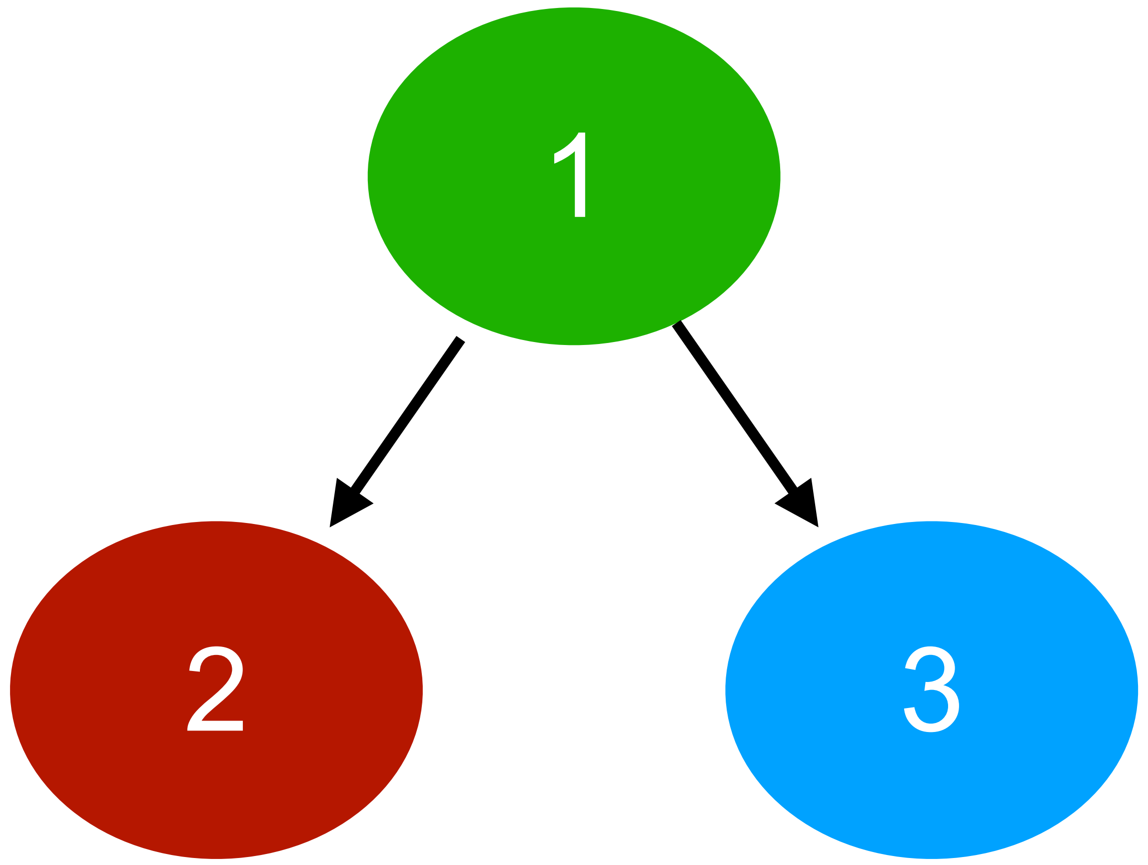
UvA - Spring 2024

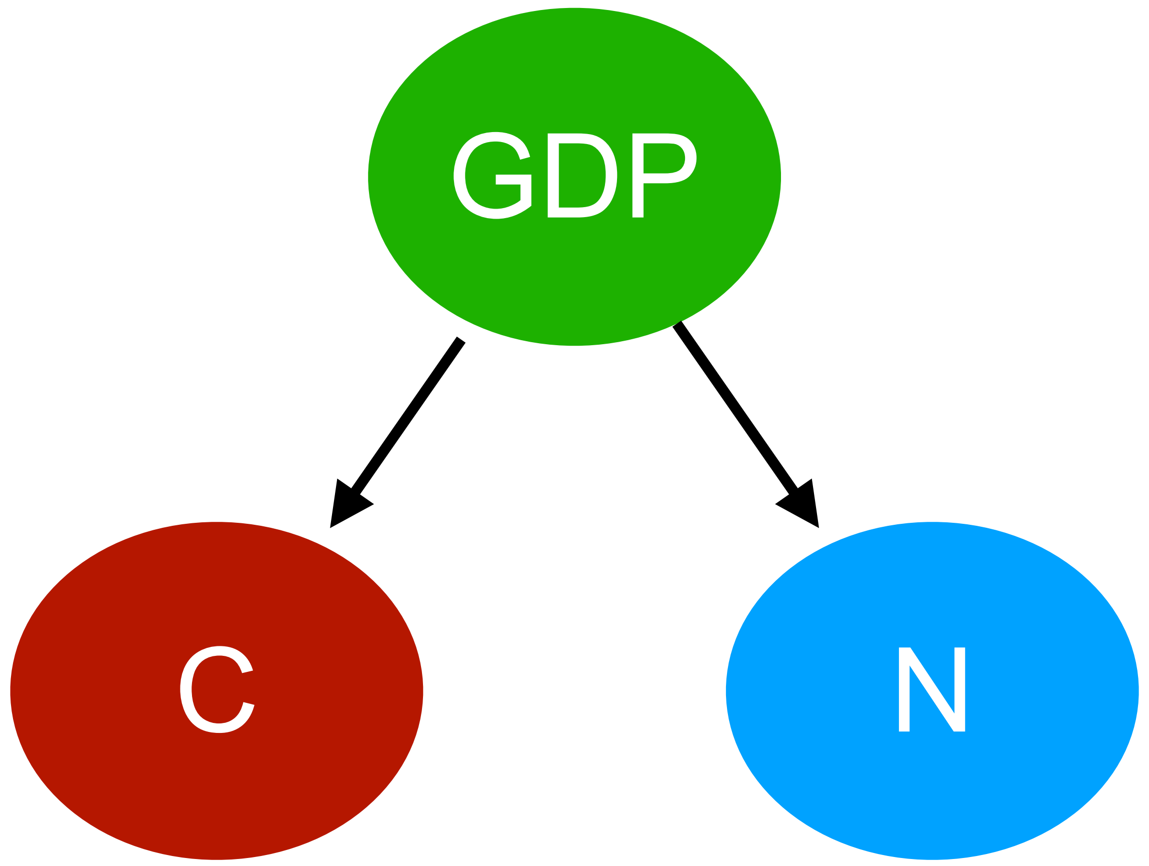
Today: graphs and random variables

- We can represent a **(factorisation of)** joint probability as a **graph**

- **Each node** $i \in \mathbf{V}$ represents a **random variable** X_i
    - For $\mathbf{A} \subseteq \mathbf{V}$, we can define $X_{\mathbf{A}} := \{X_i : i \in \mathbf{A}\}$

- **Edges** represent relationships between variables (*it will be clearer later*)





Factorizing joint distributions

- A joint distribution can always be factorized in several ways by iterating the **chain rule**, where $\mathbf{X}, \mathbf{Y}$ are disjoint sets of variables

Factorizing joint distributions

- A joint distribution can always be factorized in several ways by iterating the **chain rule**, where $\mathbf{X}, \mathbf{Y}$ are disjoint sets of variables

$$P(\mathbf{X}, \mathbf{Y}) = P(\mathbf{X} \mid \mathbf{Y})P(\mathbf{Y})$$

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Factorizing joint distributions

- A joint distribution can always be factorized in several ways by iterating the **chain rule,** where $\mathbf{X}, \mathbf{Y}$ are disjoint sets of variables

$$P(\mathbf{X}, \mathbf{Y}) = P(\mathbf{X} \mid \mathbf{Y}) P(\mathbf{Y})$$

- In general, given any **ordering** of the variables $(X_1, \ldots, X_p)$, we can write:

$$P(X_1, \ldots, X_p) = P(X_1) P(X_2 \mid X_1) \ldots P(X_p \mid X_1, \ldots, X_{p-1})$$

Equivalent factorizations

- Given any **ordering** of the variables $(X_1, \ldots, X_p)$ we can write:

$$P(X_1, \ldots, X_p) = P(X_1)P(X_2 | X_1)\ldots P(X_p | X_1, \ldots, X_{p-1})$$

- For example $P(X, Y, Z)$ can be equivalently factorized as:
    - $P(X, Y, Z) = P(X)P(Y | X)P(Z | X, Y)$
    - $P(X, Z, Y) = P(X)P(Z | X)P(Y | X, Z)$
    - $P(Z, Y, X) = P(Z)P(Y | Z)P(X | Y, Z)$ ...

These are all the same!

From probability recap: Conditional independence

- X is independent of Y **conditioned/given** Z iff

$$\forall x, y, z : P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z) \qquad (\text{for } P(Z = z) > 0)$$

- We then write $X \perp\!\!\!\perp Y \mid Z$, otherwise $X \not\perp\!\!\!\perp Y \mid Z$

From probability recap: Conditional independence

- X is independent of Y **conditioned/given** Z iff

$$\forall x, y, z : P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z) \qquad \text{(for } P(Z = z) > 0\text{)}$$

- We then write $X \perp\!\!\!\perp Y \mid Z$, otherwise $X \not\!\perp\!\!\!\perp Y \mid Z$

- Intuitively this means that **Y does not add any information** to predict X that isn't already offered by Z

- Z can be a **set of variables,** e.g. $X \perp\!\!\!\perp Y \mid \{Z_1, Z_2\}$

Example: Conditional independence

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\perp\!\!\!\perp F$$

$$P(A = 1 \mid F = 1) \neq P(A = 1 \mid F = 0) \neq P(A = 1)$$

Example: Conditional independence

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\perp\!\!\!\perp F$$

$$P(A=1 \mid F=1) \neq P(A=1 \mid F=0) \neq P(A=1)$$
$$P(A=0 \mid F=1) \neq P(A=0 \mid F=0) \neq P(A=0)$$

In this case, we just need one of these to be true to have a dependence

Example: Conditional independence

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\!\perp\!\!\!\perp F$$

$$P(A=1 \mid F=1) \neq P(A=1 \mid F=0) \neq P(A=1)$$
$$P(A=0 \mid F=1) \neq P(A=0 \mid F=0) \neq P(A=0)$$

$$A \perp\!\!\!\perp F \mid S$$

$$P(A=1 \mid F=1, S=1) = P(A=1 \mid F=0, S=1) = P(A=1 \mid S=1)$$

Example: Conditional independence

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\perp\!\!\!\perp F$$

$$P(A=1 \mid F=1) \neq P(A=1 \mid F=0) \neq P(A=1)$$
$$P(A=0 \mid F=1) \neq P(A=0 \mid F=0) \neq P(A=0)$$

$$A \perp\!\!\!\perp F \mid S$$

$$P(A=1 \mid F=1, S=1) = P(A=1 \mid F=0, S=1) = P(A=1 \mid S=1)$$

Same for other combination of values of A and S

- In this example, all information about F that A can use is contained in S

Example: Conditional independence

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\perp\!\!\!\perp F$$

$$P(A=1 \mid F=1) \neq P(A=1 \mid F=0) \neq P(A=1)$$
$$P(A=0 \mid F=1) \neq P(A=0 \mid F=0) \neq P(A=0)$$

$$A \perp\!\!\!\perp F \mid S \quad P(A=1 \mid F=1, S=1) = P(A=1 \mid F=0, S=1) = P(A=1 \mid S=1)$$

$$A \not\perp\!\!\!\perp S \mid F \quad P(A=1 \mid S=1, F=1) \neq P(A=1 \mid S=0, F=1) \neq P(A=1 \mid F=1)$$

Exploiting conditional independences

- Given any **ordering** of the variables $(X_1, \ldots, X_p)$ we can write:

$$P(X_1, \ldots, X_p) = P(X_1)P(X_2 | X_1)\ldots P(X_p | X_1, \ldots, X_{p-1})$$

- We can **simplify** the factorisation by using **conditional independences:**

$$X_i \perp\!\!\!\perp X_j | X_{\mathbf{Z}} \implies P(X_i | X_j, X_{\mathbf{Z}}) = P(X_i | X_{\mathbf{Z}})$$

(special case $X_i \perp\!\!\!\perp X_j \implies P(X_i | X_j) = P(X_i)$)

Exploiting conditional independences

- Given any **ordering** of the variables $(X_1, \ldots, X_p)$ we can write:

$$P(X_1, \ldots, X_p) = P(X_1)P(X_2 \mid X_1)\ldots P(X_p \mid X_1, \ldots, X_{p-1})$$

- We can **simplify** the factorisation by using **conditional independences:**

$$X_i \perp\!\!\!\perp X_j \mid X_{\mathbf{Z}} \implies P(X_i \mid X_j, X_{\mathbf{Z}}) = P(X_i \mid X_{\mathbf{Z}})$$

(special case $X_i \perp\!\!\!\perp X_j \implies P(X_i \mid X_j) = P(X_i)$)

Simpler factorisations mean less parameters to learn (less data)

Exploiting conditional independences

- We can **simplify** the factorisation by using **conditional independences:**

$$X_i \perp\!\!\!\perp X_j \mid X_{\mathbf{Z}} \implies P(X_i \mid X_j, X_{\mathbf{Z}}) = P(X_i \mid X_{\mathbf{Z}})$$

- For example: $X \perp\!\!\!\perp Y \mid Z$ (and symmetrically $Y \perp\!\!\!\perp X \mid Z$):
  - $P(X, Y, Z) = P(X)P(Y \mid X)P(Z \mid X, Y)$
  - $P(X, Z, Y) = P(X)P(Z \mid X)P(Y \mid X, Z)$
  - $P(Z, Y, X) = P(Z)P(Y \mid Z)P(X \mid Y, Z)$ ...

Exploiting conditional independences

- We can **simplify** the factorisation by using **conditional independences:**

$$X_i \perp\!\!\!\perp X_j \,|\, X_{\mathbf{Z}} \implies P(X_i \,|\, X_j, X_{\mathbf{Z}}) = P(X_i \,|\, X_{\mathbf{Z}})$$

- For example: $X \perp\!\!\!\perp Y \,|\, Z$ (and symmetrically $Y \perp\!\!\!\perp X \,|\, Z$):
  - $P(X, Y, Z) = P(X)P(Y \,|\, X)P(Z \,|\, X, Y)$
  - $P(X, Z, Y) = P(X)P(Z \,|\, X)P(Y \,|\, \cancel{X,} Z) = P(X)P(Z \,|\, X)P(Y \,|\, Z)$
  - $P(Z, Y, X) = P(Z)P(Y \,|\, Z)P(X \,|\, \cancel{Y,} Z) = P(Z)P(Y \,|\, Z)P(X \,|\, Z)$

Exploiting conditional independences

- We can **simplify** the factorisation by using **conditional independences:**

$$X_i \perp\!\!\!\perp X_j \mid X_{\mathbf{Z}} \implies P(X_i \mid X_j, X_{\mathbf{Z}}) = P(X_i \mid X_{\mathbf{Z}})$$

- For example: if $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z$

  - $P(X, Y, Z) = P(X)P(Y \mid X)P(Z \mid \cancel{X}, Y) = P(X)P(Y \mid X)P(Z \mid Y)$
  - $P(X, Z, Y) = P(X)P(Z \mid \cancel{X})P(Y \mid \cancel{X}, Z) = P(X)P(Z)P(Y \mid Z)$
  - $P(Z, Y, X) = P(Z)P(Y \mid Z)P(X \mid \cancel{Y}, \cancel{Z}) = P(Z)P(Y \mid Z)P(X)$

Example: Fire-smoke-alarm

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\perp\!\!\!\perp F \qquad A \perp\!\!\!\perp F \mid S \qquad P(A, F, S) \neq P(A)P(F)P(S)$$

$$P(F, S, A) = P(F)P(S \mid F)P(A \mid S, F) = P(F)P(S \mid F)P(A \mid S)$$

Example: Fire-smoke-alarm

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$A \not\perp\!\!\!\perp F \qquad A \perp\!\!\!\perp F \mid S \qquad P(A, F, S) \neq P(A)P(F)P(S)$

$$P(F, S, A) = P(F)P(S \mid F)P(A \mid S, F) = P(F)P(S \mid F)P(A \mid S)$$

$$P(A, F, S) = P(A)P(F \mid A)P(S \mid A, F)$$

Example: Fire-smoke-alarm

- A fire alarm (A) is triggered by smoke (S), which is usually caused by fire (F)

$$A \not\perp\!\!\!\perp F \quad A \perp\!\!\!\perp F \mid S \qquad P(A, F, S) \neq P(A)P(F)P(S)$$

$$P(F, S, A) = P(F)P(S \mid F)P(A \mid S, F) = P(F)P(S \mid F)P(A \mid S)$$

$$P(A, F, S) = P(A)P(F \mid A)P(S \mid A, F)$$

$$P(S, F, A) = P(S)P(F \mid S)P(A \mid S, F) = P(S)P(F \mid S)P(A \mid S)$$

Bayesian networks

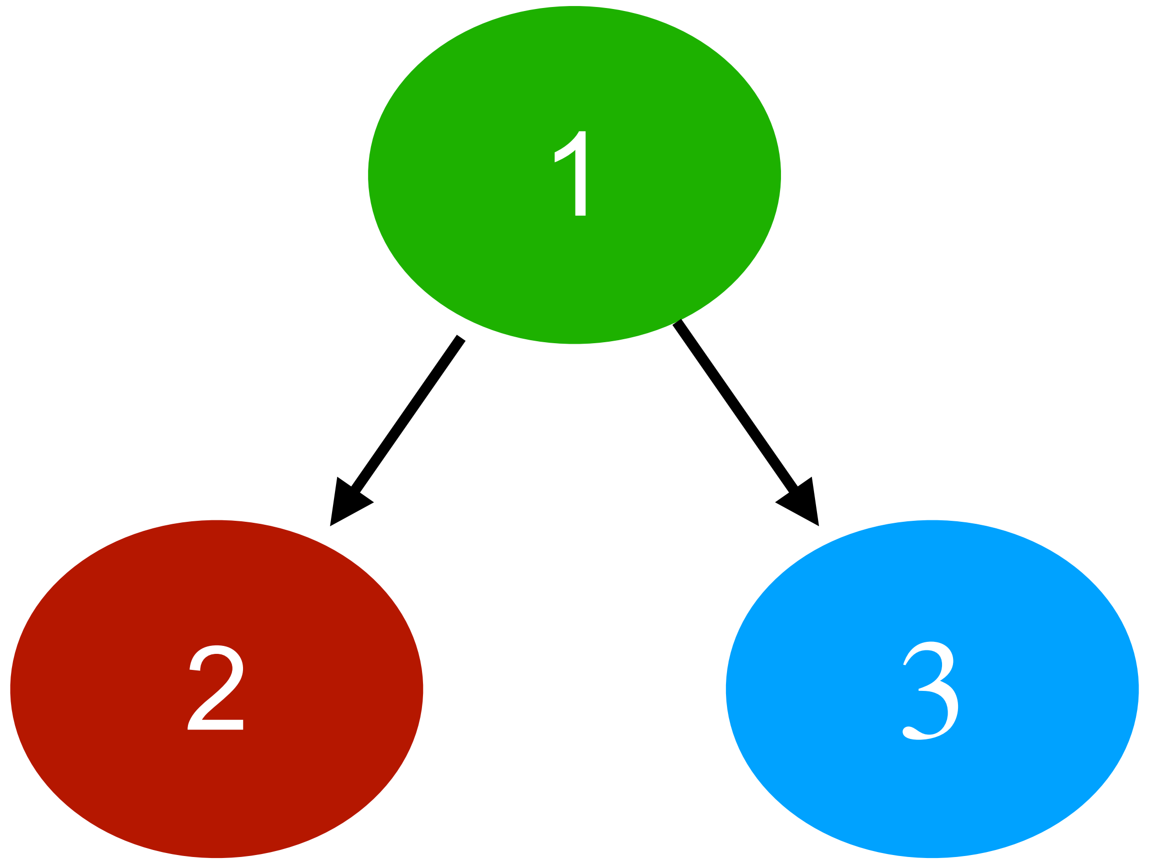
- We have a set of random variables $X_1, \ldots, X_p$ with joint $p(X_1, \ldots, X_p)$
- We have a DAG G, s.t. **each random variable** X_i is represented by **node** i

- We then say $p(X_1, \ldots, X_p)$ **factorizes over G** if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \,|\, \mathbf{X}_{\mathrm{Pa}(i)})$$

- A **Bayesian network** (BN) is the tuple (G, p) s.t. p **factorizes over** G

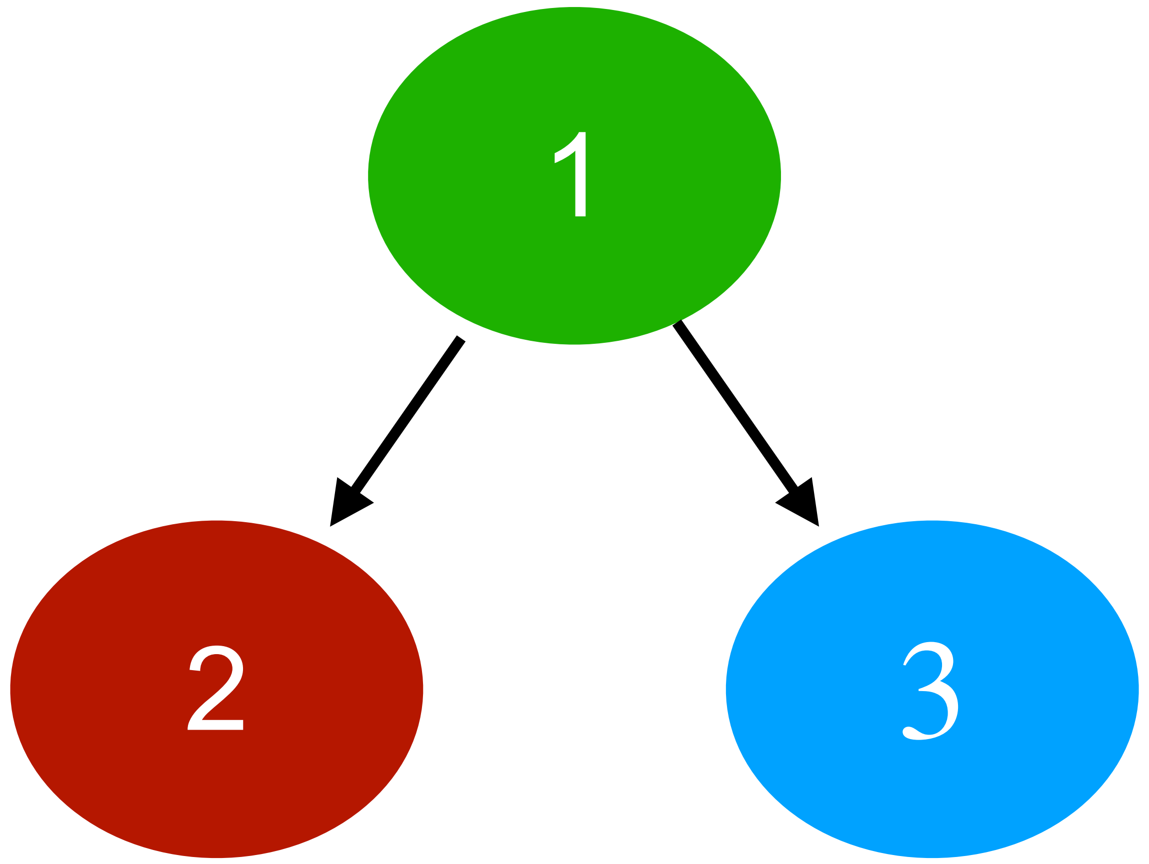
Example Bayesian networks



$$p(X_1, X_2, X_3) = p(X_1 \mid X_{\mathrm{Pa}(1)}) p(X_2 \mid X_{\mathrm{Pa}(2)}) p(X_3 \mid X_{\mathrm{Pa}(3)})$$

$$p(X_1) \quad p(X_2 \mid X_1) \quad p(X_3 \mid X_1)$$

Example Bayesian networks

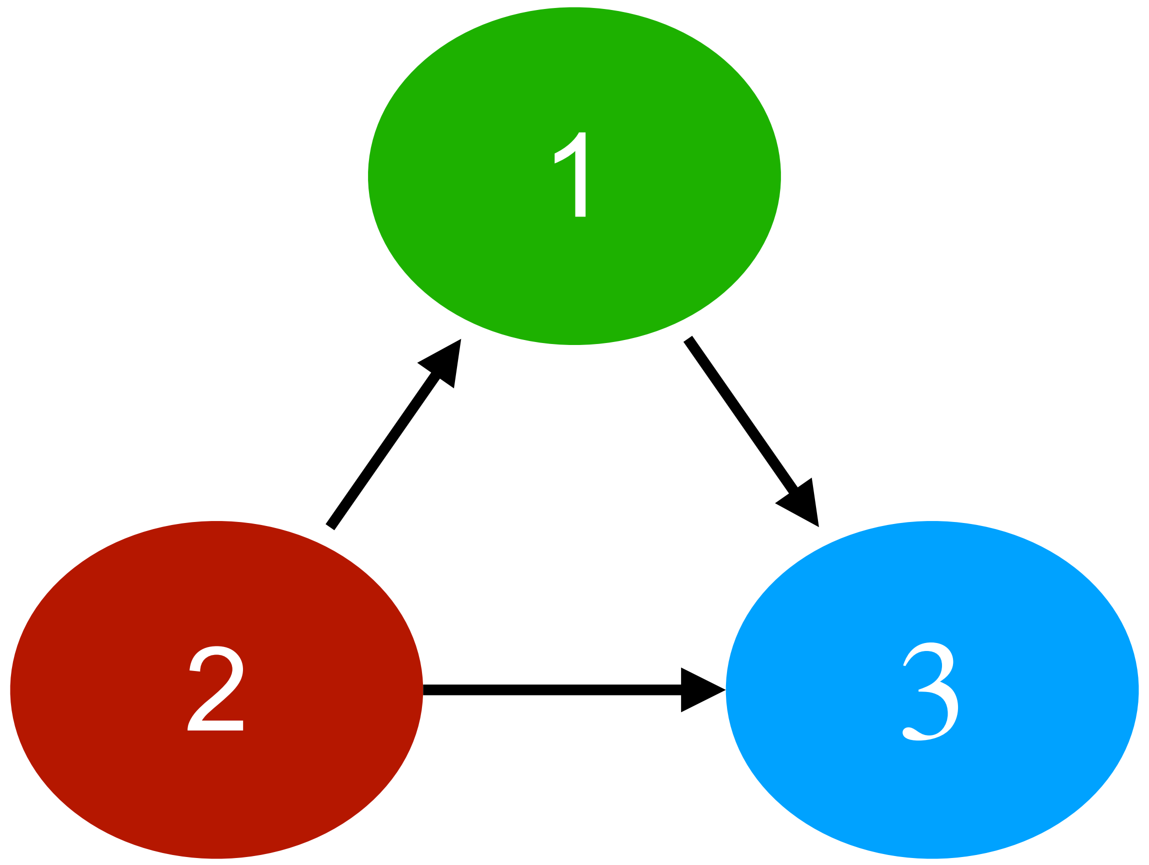


$$p(X_1, X_2, X_3) = p(X_1 \mid X_{\mathrm{Pa}(1)})p(X_2 \mid X_{\mathrm{Pa}(2)})p(X_3 \mid X_{\mathrm{Pa}(3)})$$

$$p(X_1) \quad p(X_2 \mid X_1) \quad p(X_3 \mid X_1)$$

The DAG/factorization is not unique for a given joint distribution:

(for example any fully connected graph factorizes any p)



$$p(X_1, X_2, X_3) = p(X_2)p(X_1 \mid X_2)p(X_3 \mid X_1, X_2)$$

Equivalent factorizations

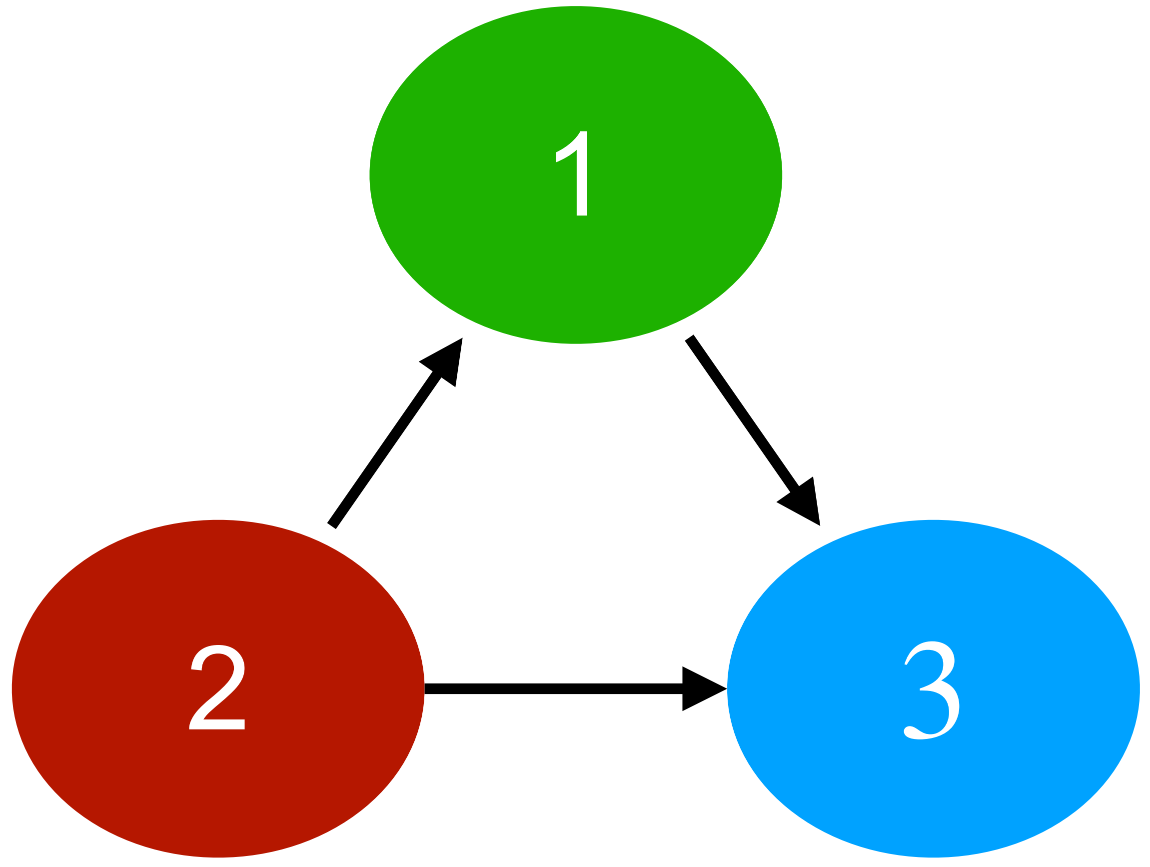
- Given any **ordering** of the variables $(X_1, \ldots, X_p)$ we can write:

$$P(X_1, \ldots, X_p) = P(X_1)P(X_2 | X_1)\ldots P(X_p | X_1, \ldots, X_{p-1})$$

- For example $P(X, Y, Z)$ can be equivalently factorized as:
  - $P(X, Y, Z) = P(X)P(Y | X)P(Z | X, Y)$
  - $P(X, Z, Y) = P(X)P(Z | X)P(Y | X, Z)$
  - $P(Z, Y, X) = P(Z)P(Y | Z)P(X | Y, Z)$ ...

Each of these factorizes according to a fully connected DAG

Example Bayesian networks



If we know $X_3 \perp\!\!\!\perp X_2 \mid X_1$ then we can simplify the factorisation

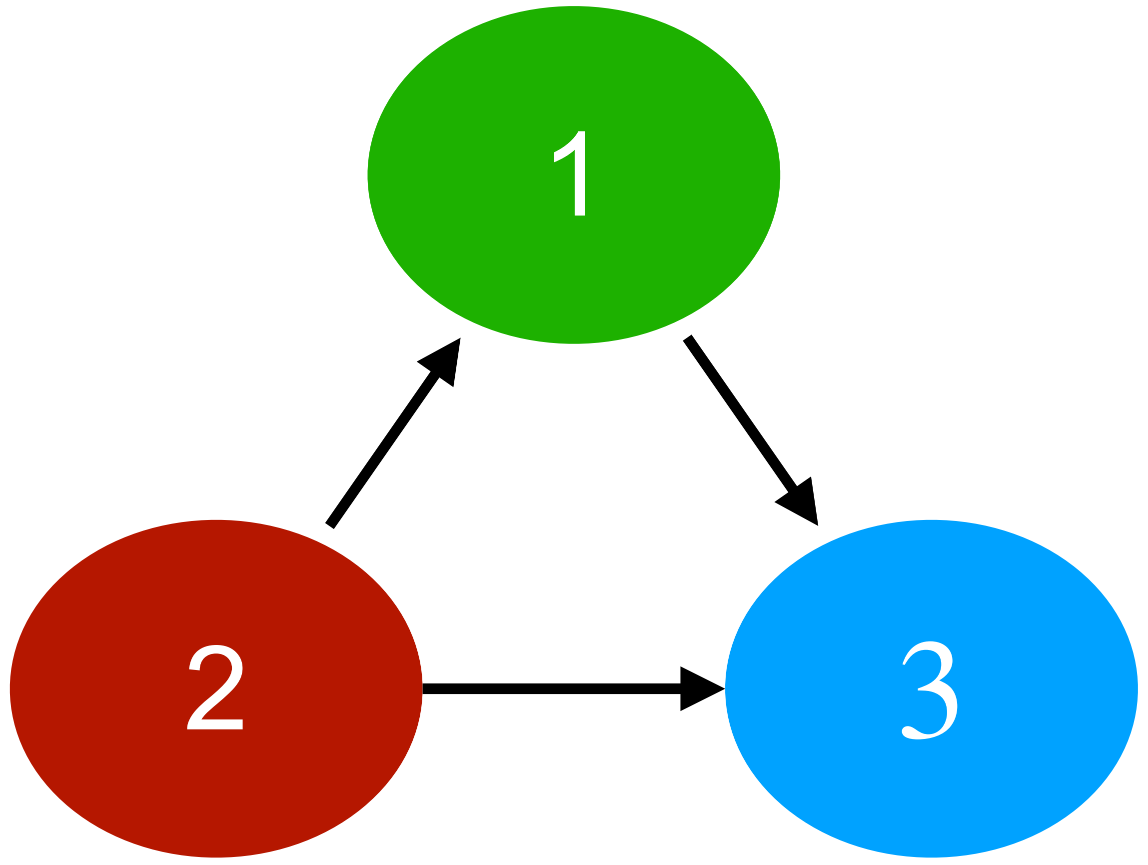
$$p(X_1, X_2, X_3) = p(X_2)p(X_1 \mid X_2)p(X_3 \mid X_1, \cancel{X_2})$$

Note: The fully connected graph still ***factorizes*** according p, even if we can simplify factorisation, since this still holds

$$p(X_1 \mid X_{\mathrm{Pa}(1)})p(X_2 \mid X_{\mathrm{Pa}(2)})p(X_3 \mid X_{\mathrm{Pa}(3)})$$

$$p(X_1 \mid X_2)p(X_2)p(X_3 \mid X_2, X_1) = p(X_1, X_2, X_3)$$

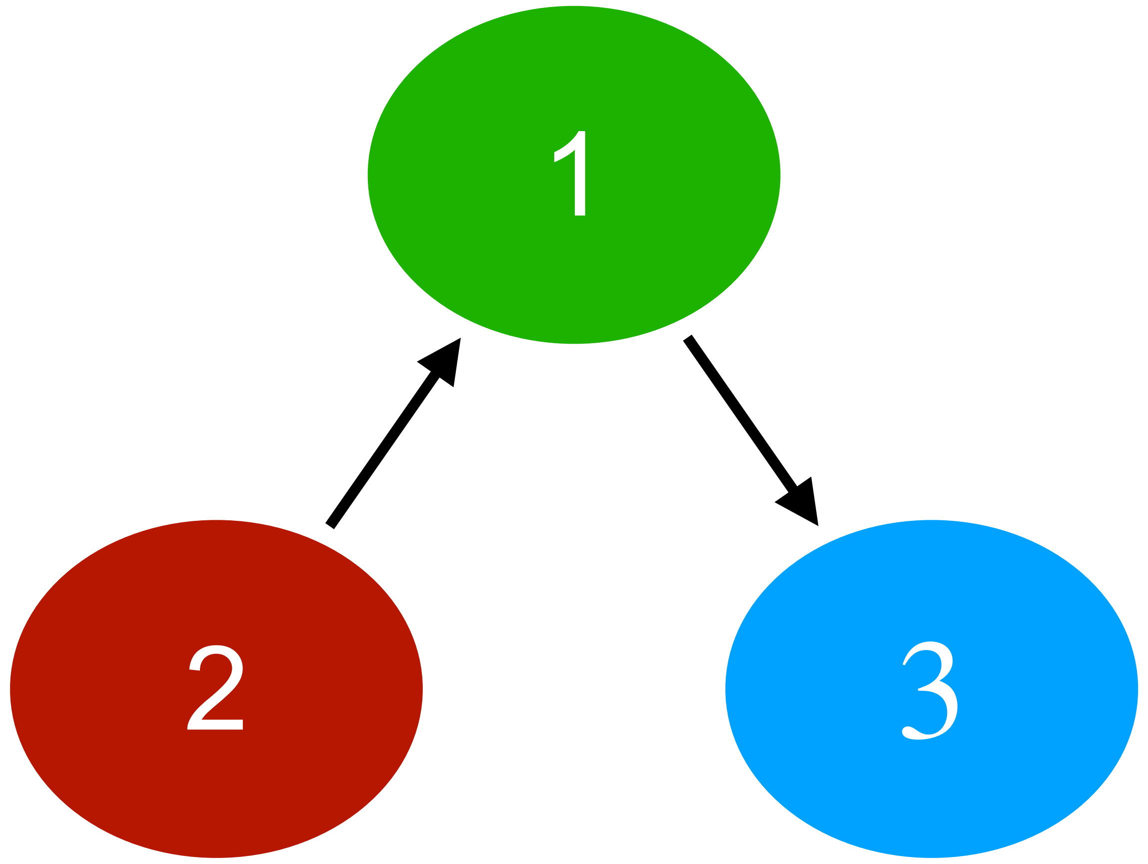
Example Bayesian networks



If we know $X_3 \perp\!\!\!\perp X_2 \,|\, X_1$ then we can simplify the factorisation

$$p(X_1, X_2, X_3) = p(X_2)p(X_1 \,|\, X_2)p(X_3 \,|\, X_1, \cancel{X_2})$$

Note: The fully connected graph still ***factorizes*** according p, even if we can simplify factorisation, since this still holds



But if $X_3 \perp\!\!\!\perp X_2 \,|\, X_1$ holds in p, then this DAG also factorizes accordingly

$$p(X_1 \,|\, X_{\mathrm{Pa}(1)})p(X_2 \,|\, X_{\mathrm{Pa}(2)})p(X_3 \,|\, X_{\mathrm{Pa}(3)})$$

$$p(X_1 \,|\, X_2) \qquad p(X_2) \qquad p(X_3 \,|\, X_1)$$

Why should we care about Bayesian networks?

- We have a set of random variables $X_1, \ldots, X_p$ with joint $p(X_1, \ldots, X_p)$
- We have a DAG G, s.t. **each random variable** X_i is represented by **node** i

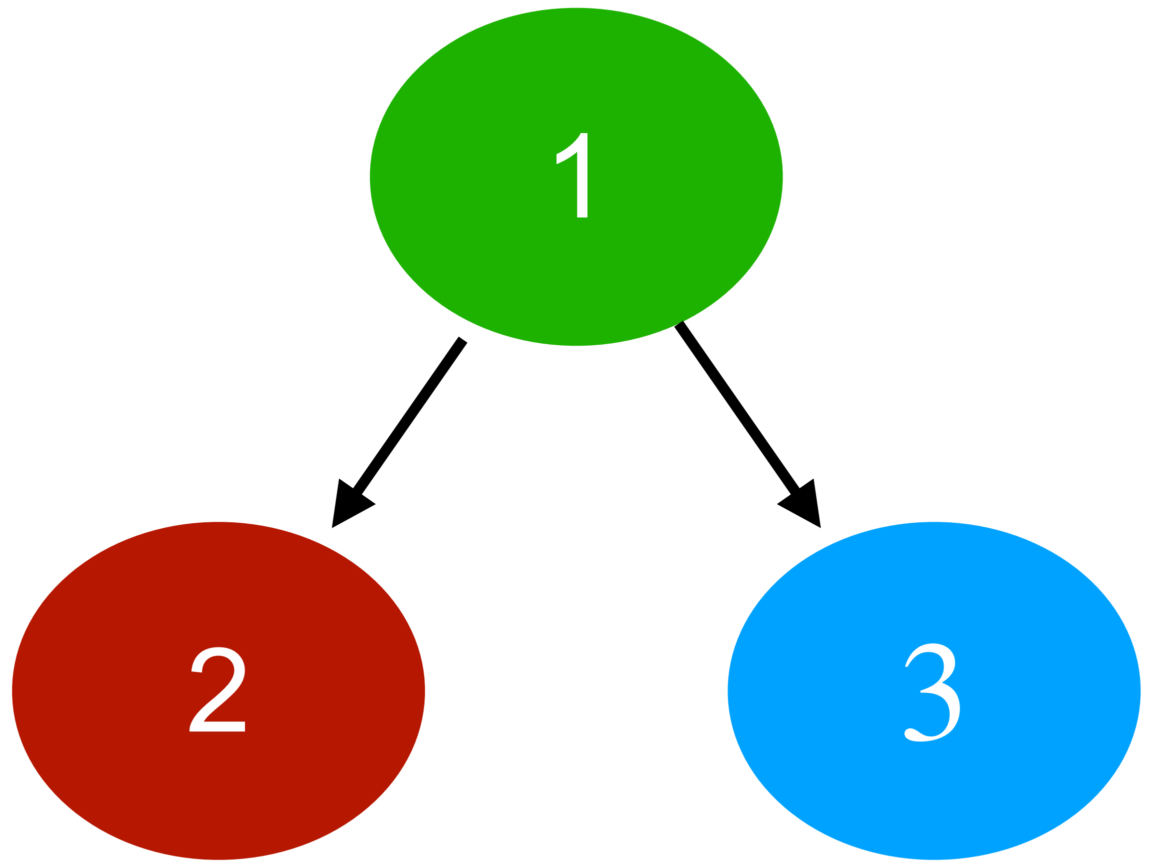
- We then say $p(X_1, \ldots, X_p)$ **factorizes over G** if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

They can help simplify the factorisation

We can easily read conditional independences

They can represent causal models



Global Markov Property and faithfulness

- If (G, p) is a Bayesian network with a DAG $G = (\mathbf{V}, \mathbf{E})$, i.e. **p factorizes according to G**, then for any disjoint $\mathbf{A}, \mathbf{B}, \mathbf{C} \subseteq \mathbf{V}$:

$$\mathbf{A} \perp_d \mathbf{B} \,|\, \mathbf{C} \implies X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \,|\, X_{\mathbf{C}}$$

Global Markov Property and faithfulness

- If (G, p) is a Bayesian network with a DAG $G = (\mathbf{V}, \mathbf{E})$, i.e. **p factorizes according to G**, then for any disjoint $\mathbf{A}, \mathbf{B}, \mathbf{C} \subseteq \mathbf{V}$:

$$\mathbf{A} \perp_d \mathbf{B} \,|\, \mathbf{C} \implies X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \,|\, X_{\mathbf{C}}$$

- **d-separations** that can be read purely from a graph imply **conditional independences** in the random variables and data generated by the graph

Global Markov Property and faithfulness

- If (G, p) is a Bayesian network with a DAG $G = (\mathbf{V}, \mathbf{E})$, i.e. **p factorizes according to G**, then for any disjoint $\mathbf{A}, \mathbf{B}, \mathbf{C} \subseteq \mathbf{V}$:

$$\mathbf{A} \perp_d \mathbf{B} \,|\, \mathbf{C} \implies X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \,|\, X_{\mathbf{C}}$$

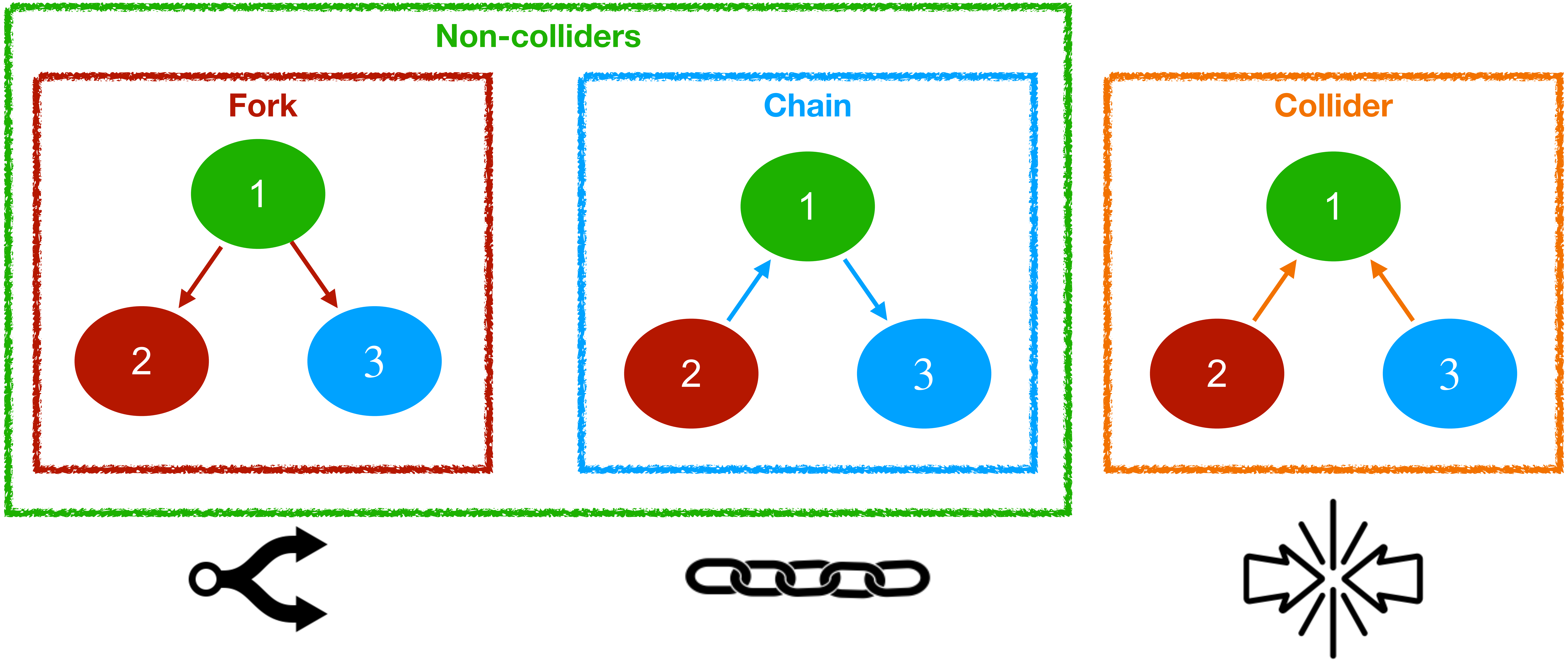
- **d-separations** that can be read purely from a graph imply **conditional independences** in the random variables and data generated by the graph

- The reverse implication is not true in general, but if it is **p is faithful to G**

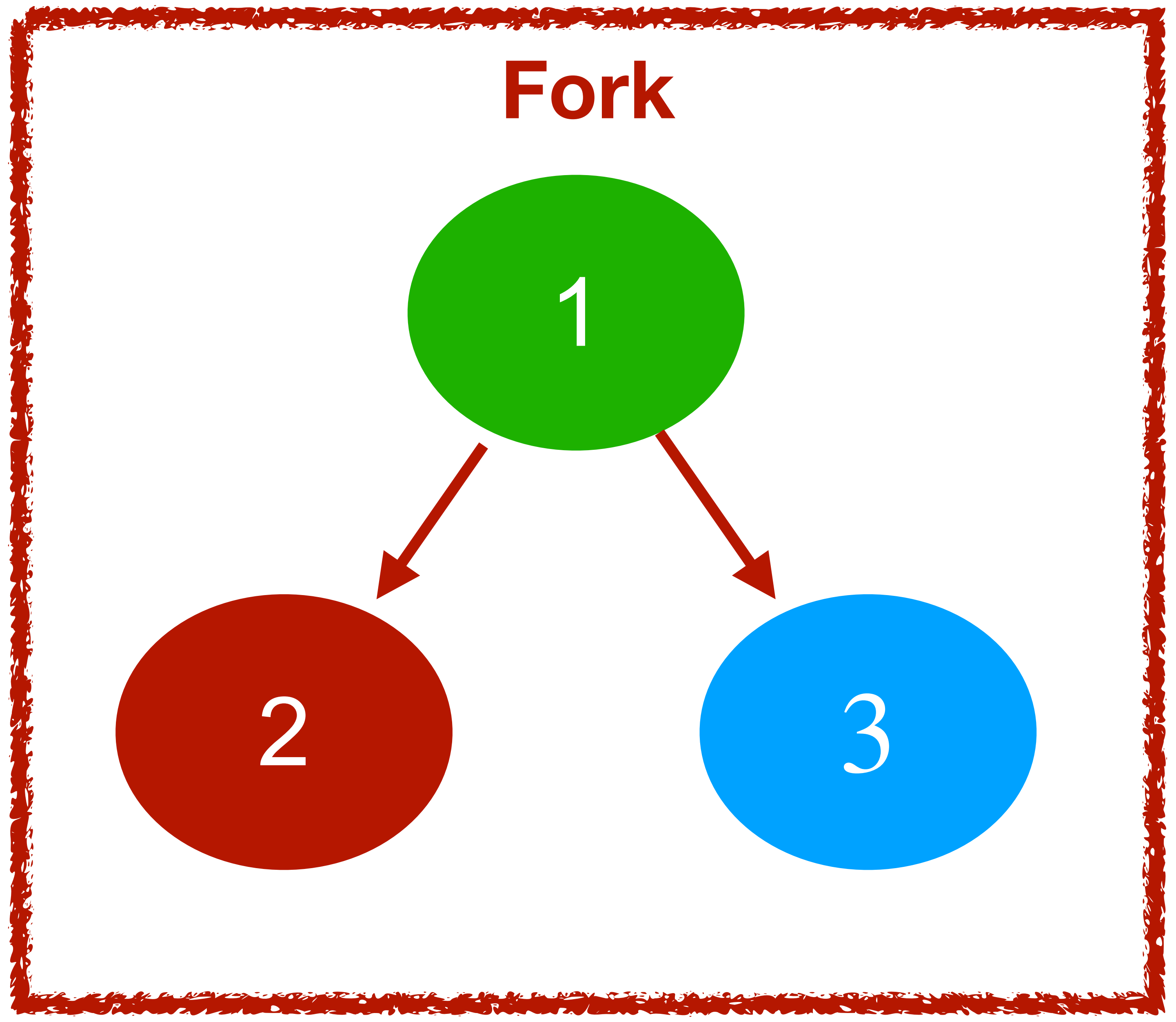
$$\mathbf{A} \perp_d \mathbf{B} \,|\, \mathbf{C} \iff X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \,|\, X_{\mathbf{C}}$$

We will usually assume both assumptions hold

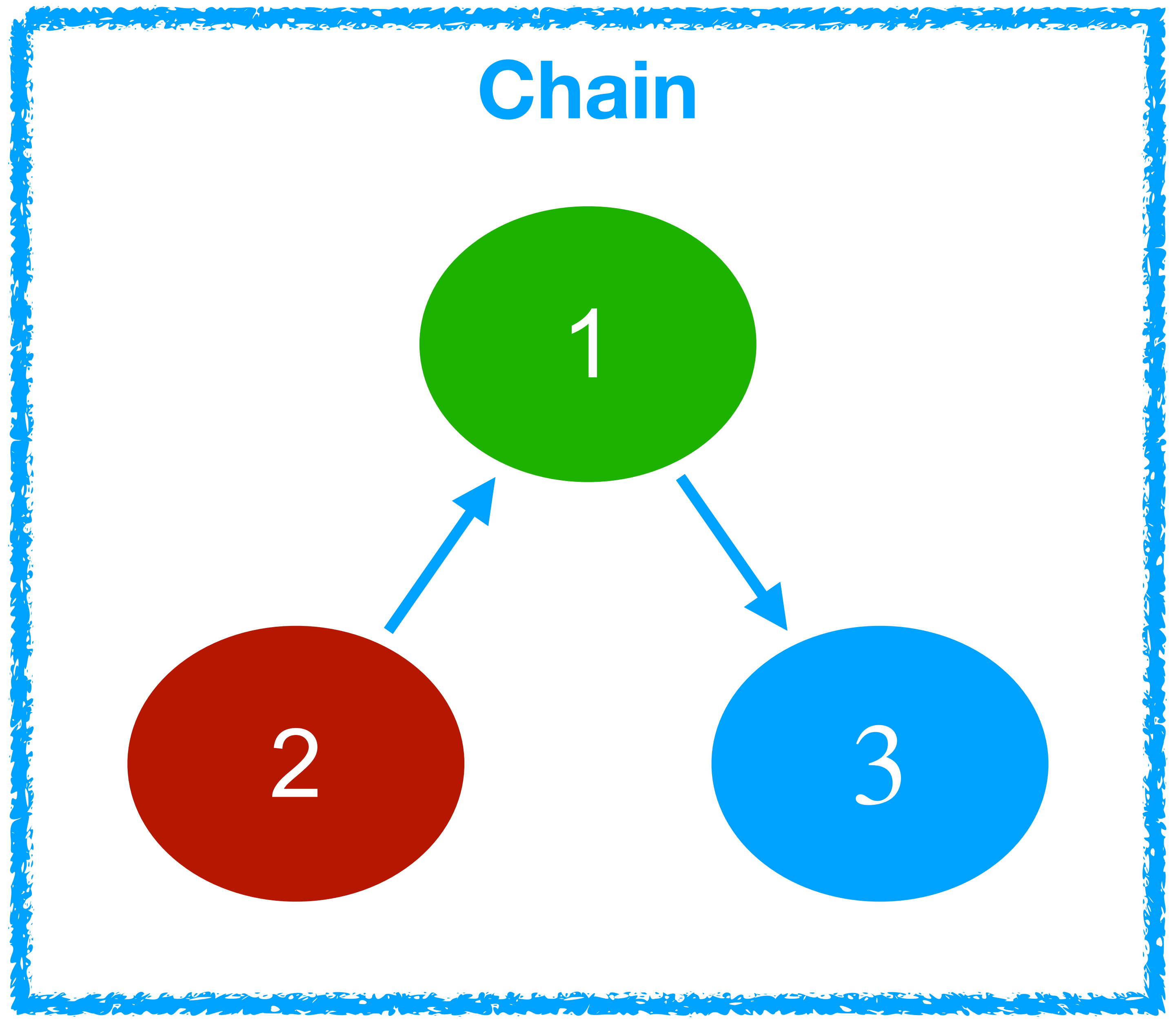
D-separation (reprise): three basic patterns



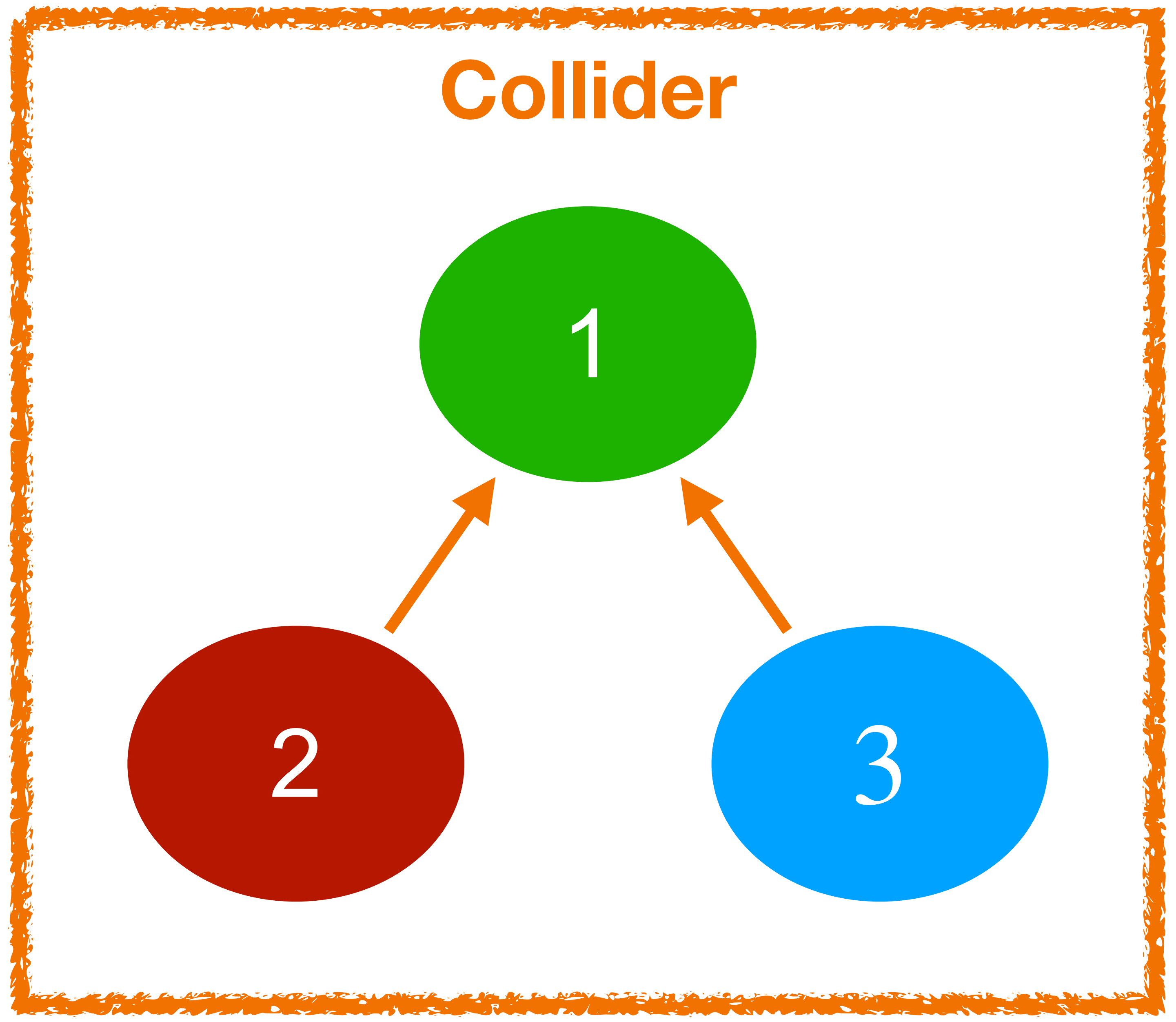
D-separation (reprise): three basic patterns



$$P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)$$

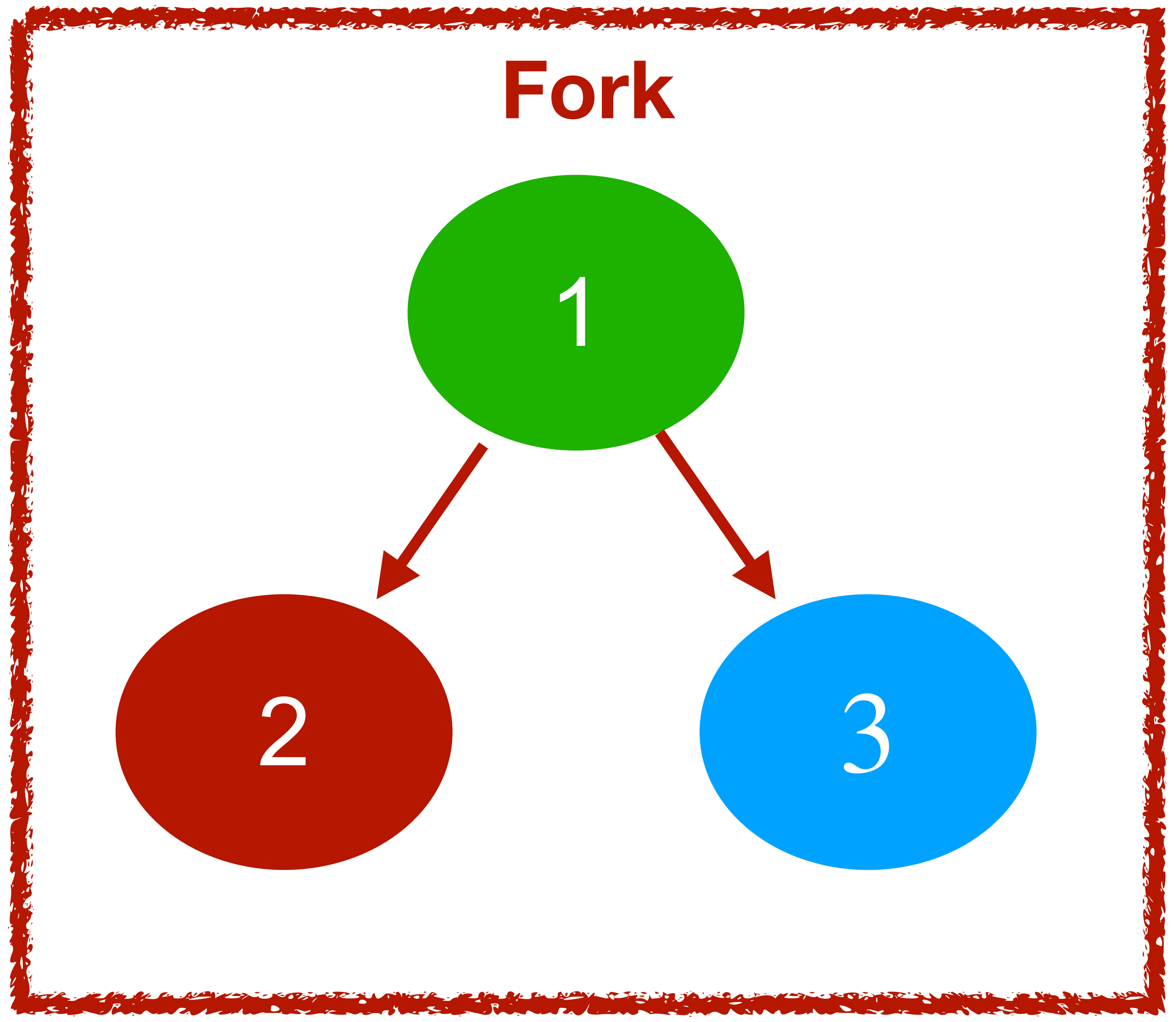


$$P(X_2)P(X_1 \mid X_2)P(X_3 \mid X_1)$$



$$P(X_2)P(X_3)P(X_1 \mid X_2, X_3)$$

Fork (shoe size <- age -> height)



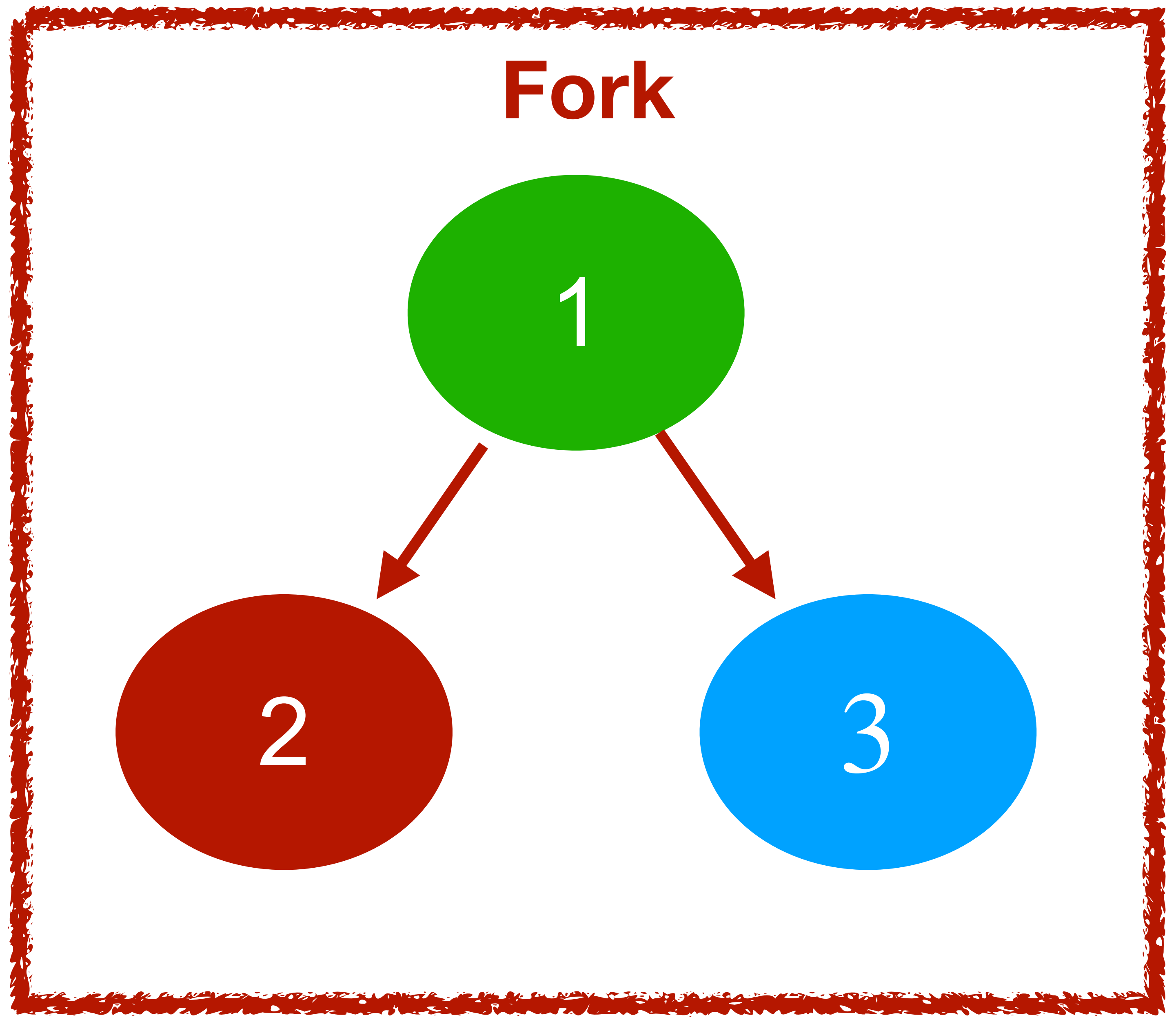
$$P(X_1)P(X_2|X_1)P(X_3|X_1)$$

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)$$

$$3 \perp_d 2|1 \implies X_3 \perp\!\!\!\perp X_2|X_1$$

$$= P(X_1)P(X_2|X_1)P(X_3|X_1)$$

Fork (shoe size <- age -> height)



$$P(X_1)P(X_2 | X_1)P(X_3 | X_1)$$

$$P(X_1, X_2, X_3) = P(X_1)P(X_2 | X_1)P(X_3 | X_1, X_2)$$

$$3 \perp_d 2 | 1 \implies X_3 \perp\!\!\!\perp X_2 | X_1$$

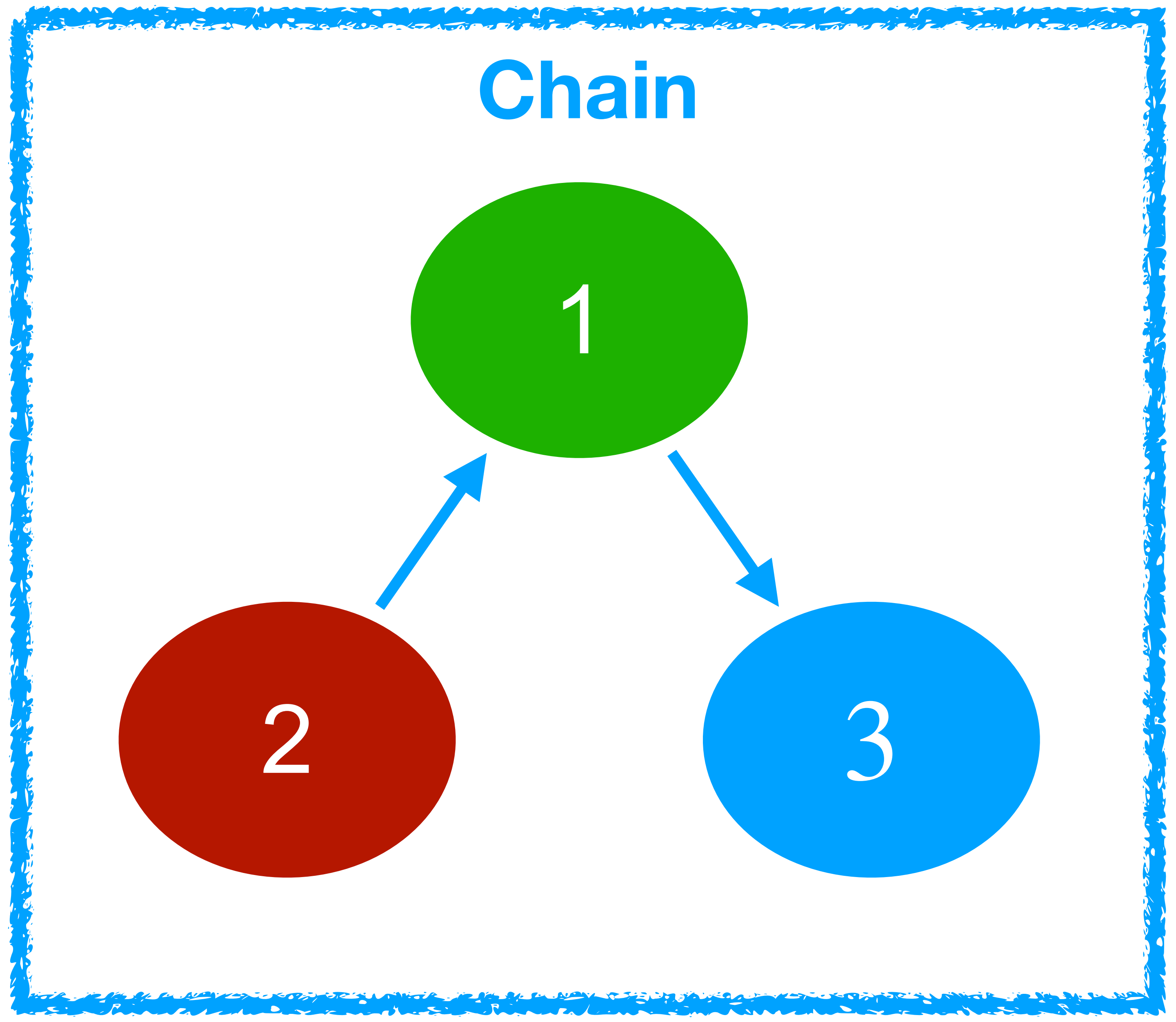
$$= P(X_1)P(X_2 | X_1)P(X_3 | X_1)$$

In general $2 \not\perp_d 3 \not\Longrightarrow X_2 \not\perp\!\!\!\perp X_3$, because

$$\mathbf{A} \perp_d \mathbf{B} | \mathbf{C} \implies X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} | X_{\mathbf{C}}$$

In practice, for faithful graphs $2 \not\perp_d 3 \iff X_2 \not\perp\!\!\!\perp X_3$

Chain (fire -> smoke -> alarm)



$$P(X_2)P(X_1 \mid X_2)P(X_3 \mid X_1)$$

$$P(X_1, X_2, X_3) = P(X_2)P(X_1 \mid X_2)P(X_3 \mid X_1, X_2)$$

$$3 \perp_d 2 \mid 1 \implies X_3 \perp\!\!\!\perp X_2 \mid X_1$$

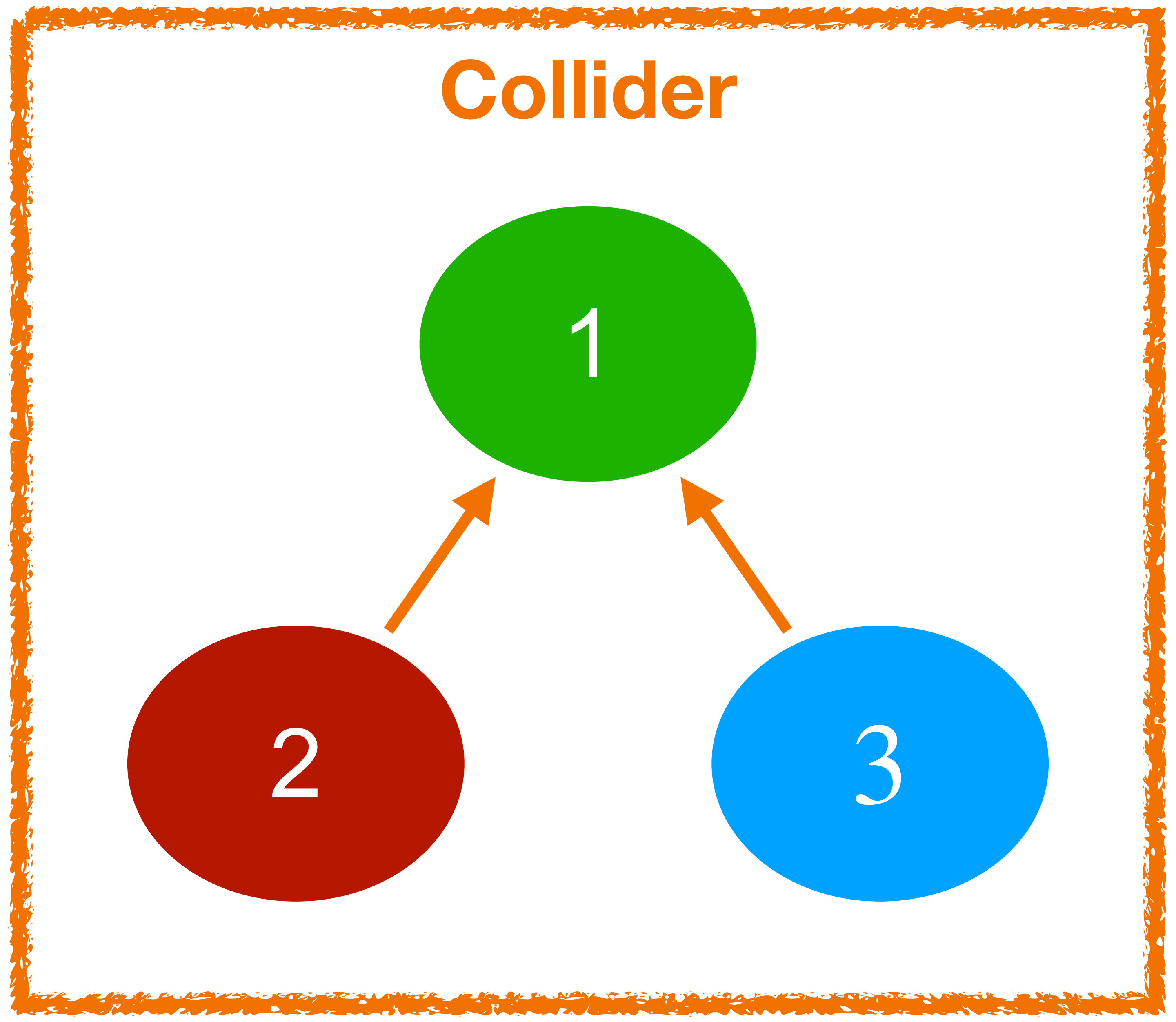
$$= P(X_2)P(X_1 \mid X_2)P(X_3 \mid X_1)$$

In general $2 \not\perp_d 3 \not\Longrightarrow X_2 \not\!\perp\!\!\!\perp X_3$, because

$$\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C} \implies X_\mathbf{A} \perp\!\!\!\perp X_\mathbf{B} \mid X_\mathbf{C}$$

In practice, for faithful graphs $2 \not\perp_d 3 \iff X_2 \not\!\perp\!\!\!\perp X_3$

Collider (Math skills -> admission <- English skills)



$P(X_2)P(X_3)P(X_1 \mid X_2, X_3)$

$$P(X_1, X_2, X_3) = P(X_2)P(X_3 \mid X_2)P(X_1 \mid X_2, X_3)$$

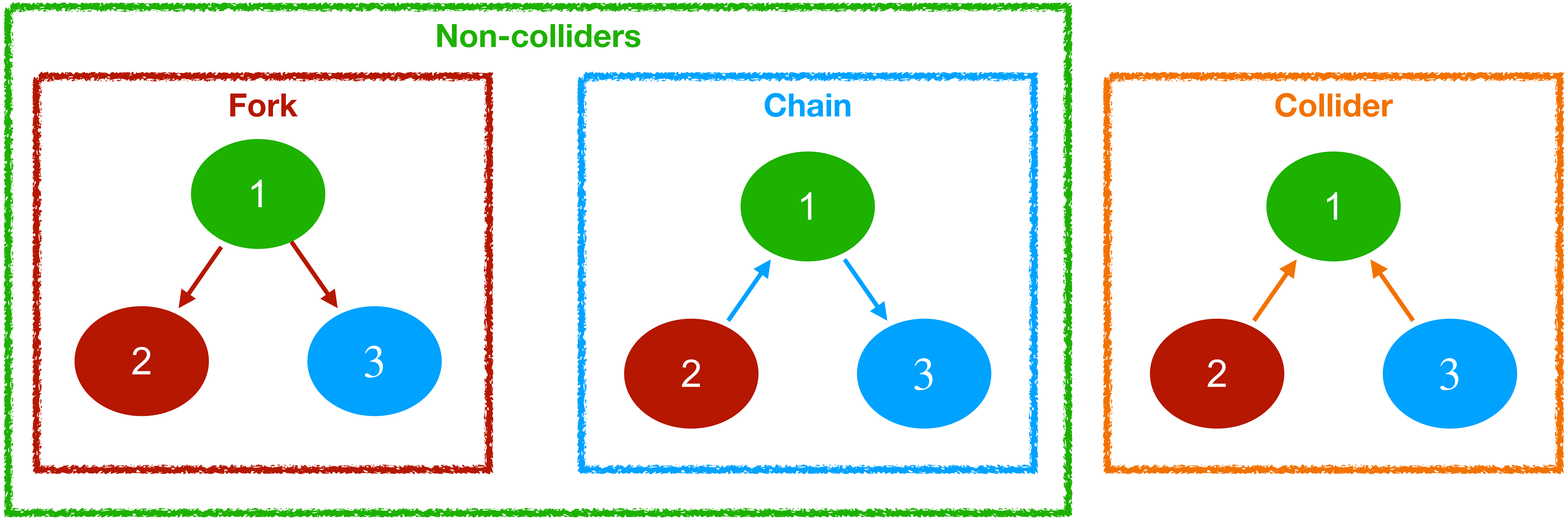
$$3 \perp_d 2 \implies X_3 \perp\!\!\!\perp X_2$$

$$= P(X_2)P(X_3)P(X_1 \mid X_2, X_3)$$

In general $2 \not\perp_d 3 \mid 1 \;\not\Longrightarrow\; X_2 \not\perp\!\!\!\perp X_3 \mid X_1$

For faithful graphs $2 \not\perp_d 3 \mid 1 \iff X_2 \not\perp\!\!\!\perp X_3 \mid X_1$

Three basic patterns



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

$X_3 \perp\!\!\!\perp X_2$

Why should we care about Bayesian networks?

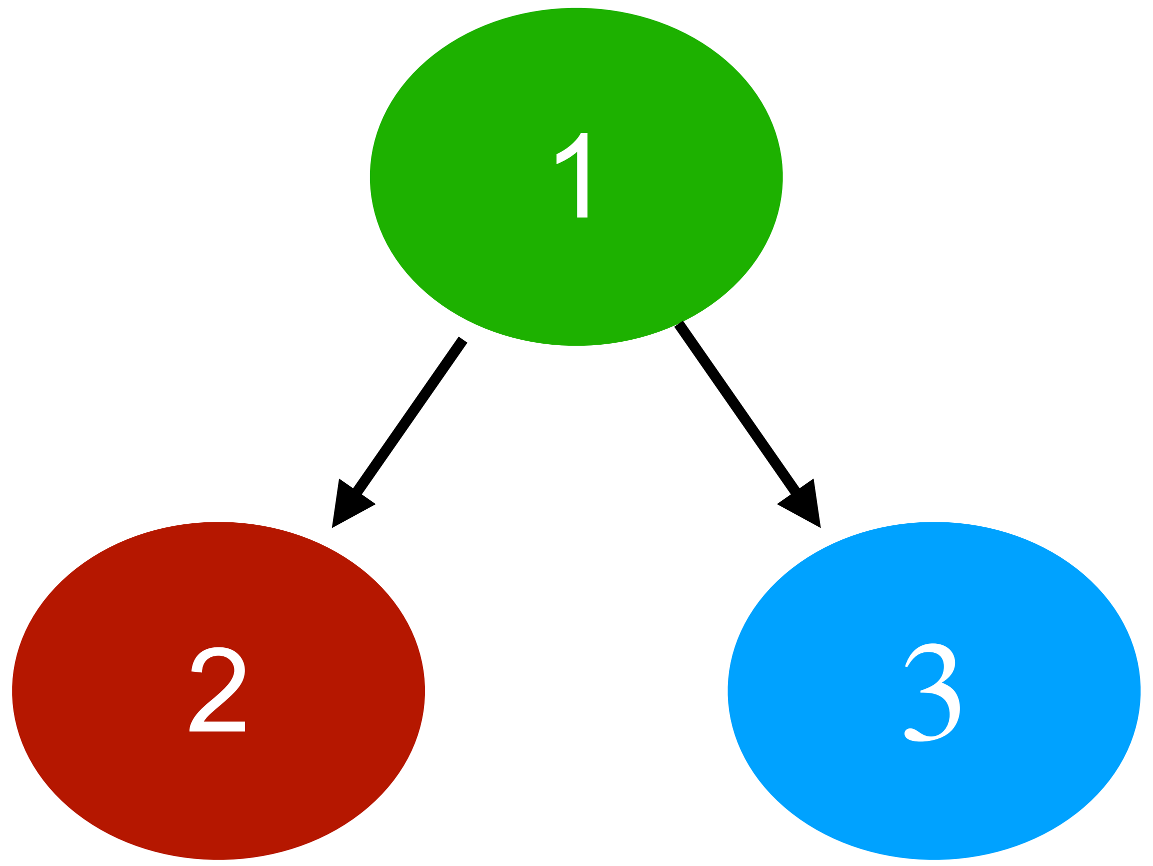
- We have a set of random variables $X_1, \ldots, X_p$ with joint $p(X_1, \ldots, X_p)$
- We have a DAG G, s.t. **each random variable** X_i is represented by **node** i
- We then say $p(X_1, \ldots, X_p)$ **factorizes over G** if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \,|\, \mathbf{X}_{\mathrm{pa}(i)})$$

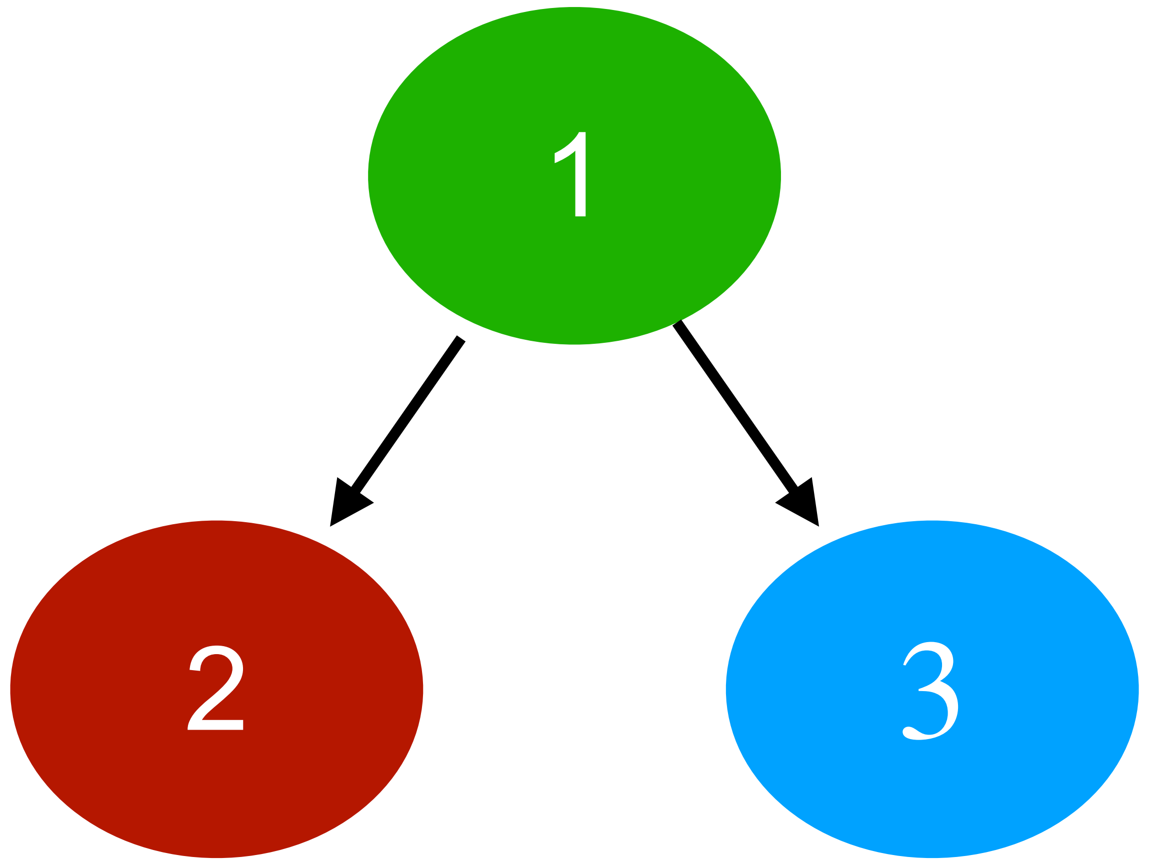
They can help simplify the factorisation

We can easily read conditional independences

They can represent causal models



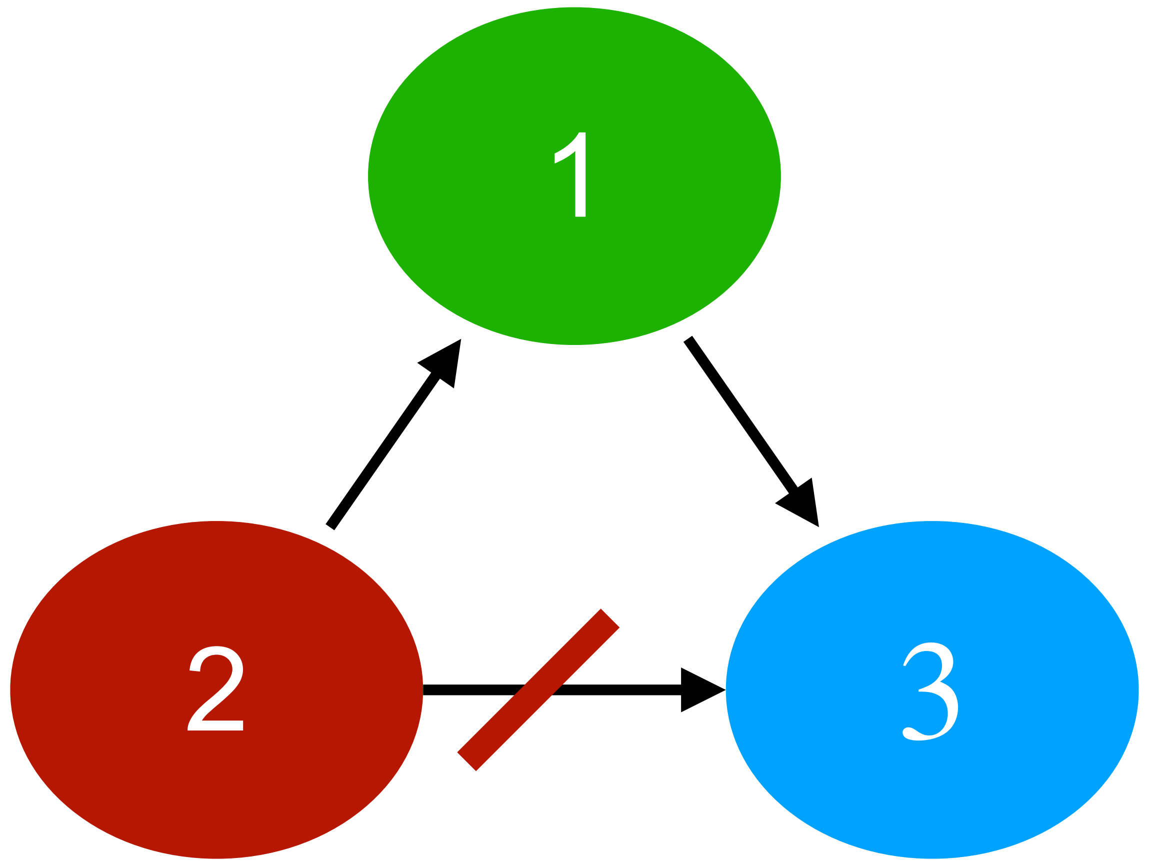
Example Bayesian networks



$$p(X_1, X_2, X_3) = p(X_1 \mid X_{\mathrm{Pa}(1)})p(X_2 \mid X_{\mathrm{Pa}(2)})p(X_3 \mid X_{\mathrm{Pa}(3)})$$

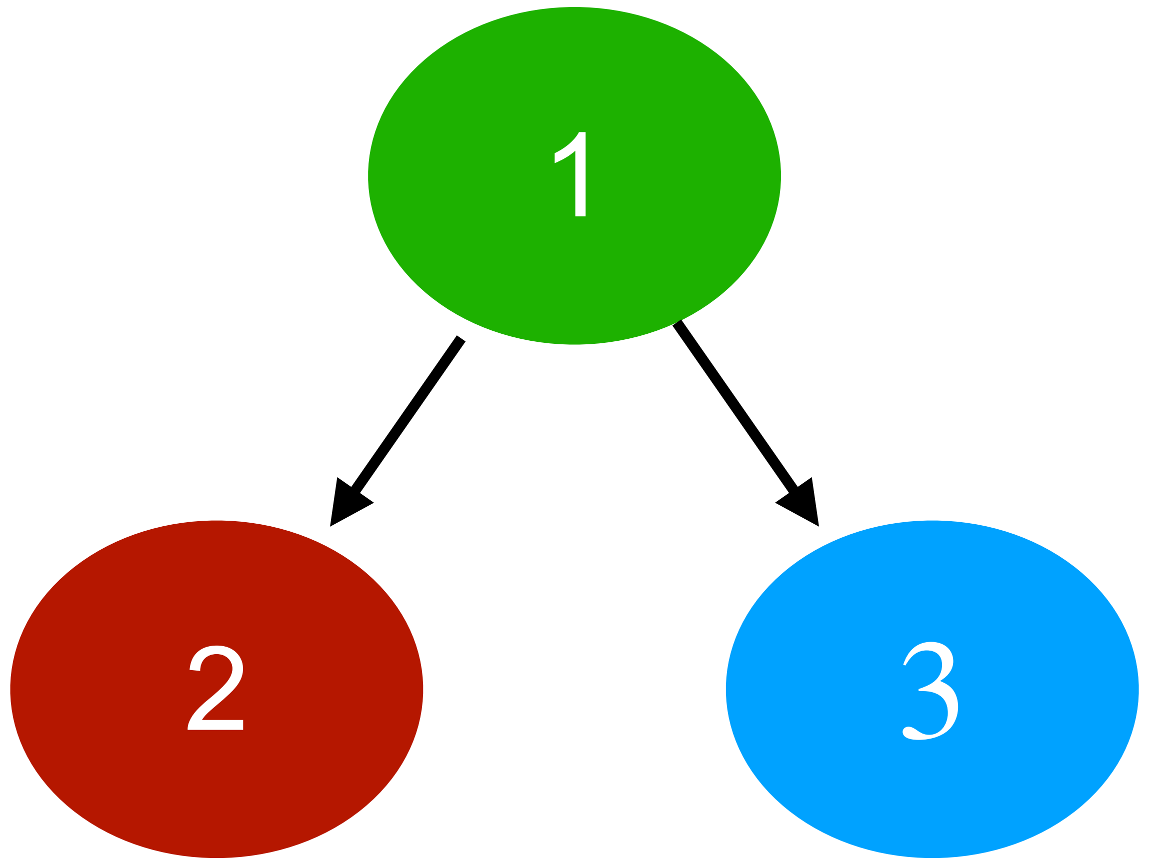
$$p(X_1) \quad p(X_2 \mid X_1) \quad p(X_3 \mid X_1)$$

$$3 \perp_d 2 \mid 1 \implies X_3 \perp\!\!\!\perp X_2 \mid X_1$$



$$p(X_1, X_2, X_3) = p(X_2)p(X_1 \mid X_2)p(X_3 \mid X_1, \cancel{X_2})$$

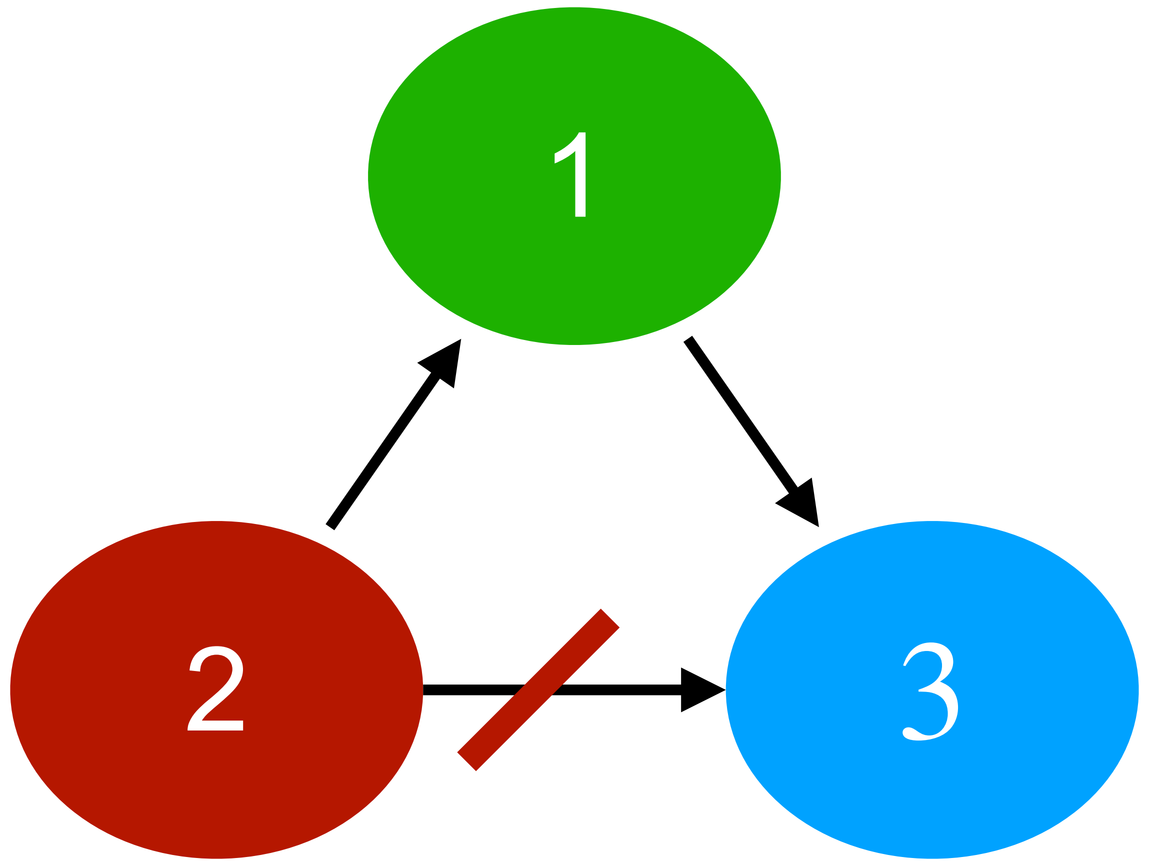
Example Bayesian networks



$$p(X_1, X_2, X_3) = p(X_1 \mid X_{\mathrm{Pa}(1)}) p(X_2 \mid X_{\mathrm{Pa}(2)}) p(X_3 \mid X_{\mathrm{Pa}(3)})$$

$$p(X_1) \quad p(X_2 \mid X_1) \quad p(X_3 \mid X_1)$$

$$3 \perp_d 2 \mid 1 \implies X_3 \perp\!\!\!\perp X_2 \mid X_1$$



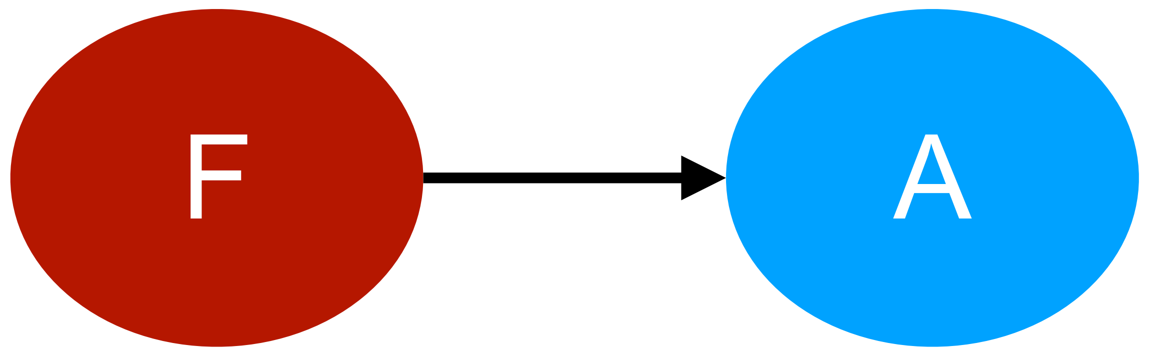
$$p(X_1, X_2, X_3) = p(X_2) p(X_1 \mid X_2) p(X_3 \mid X_1, \cancel{X_2})$$

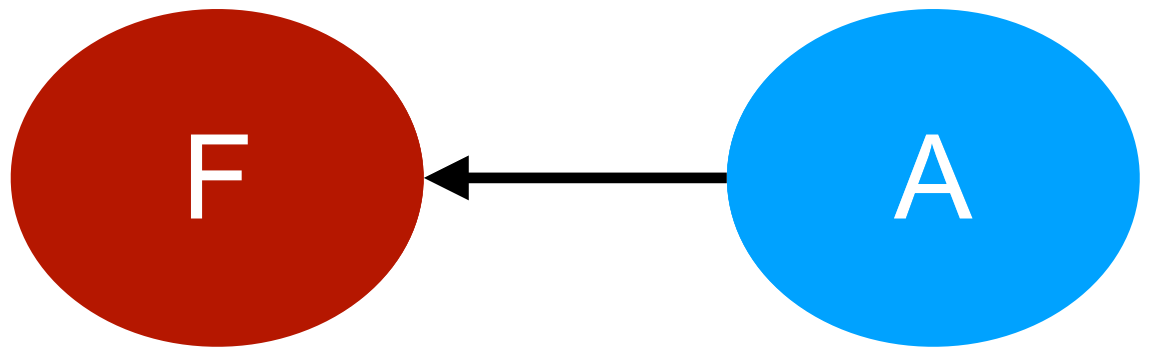
We have multiple "minimal" graphs for a given distribution, which is the causal one?

Next time: BNs vs causal BNs

- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$$p(F, A) = p(F)\, p(A|F) \qquad\qquad p(F, A) = p(A)\, p(F|A)$$





Next time: BNs vs causal BNs

- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$$p(F, A) = p(F)\, p(A \mid F) \qquad\qquad p(F, A) = p(A)\, p(F \mid A)$$

$$F \rightarrow A \qquad\qquad F \leftarrow A$$

CAUSAL (lighting a fire triggers alarm)

NOT-CAUSAL (triggering alarm does not light a fire)

Next time: BNs vs causal BNs

1	Introduction
2	Probability recap
3	Graphical models, d-separation
4	Causal graphs, Interventions, SCMs
5	Covariate adjustment: backdoor criterion
6	Covariate Frontdoor criterion, Instrumental variables
7	Counterfactuals, potential outcomes, estimating causal effects 1
8	Estimating causal effects 2 (matching, IPW)
9	Constraint based structure learning
10	Score based structure learning, restricted models
11	Do-calculus, transportability, Joint Causal Inference
12	Causality-inspired ML, recap of the course

For more details see: https://canvas.uva.nl/courses/36213/pages/course-schedule-tentative?module_item_id=1585152

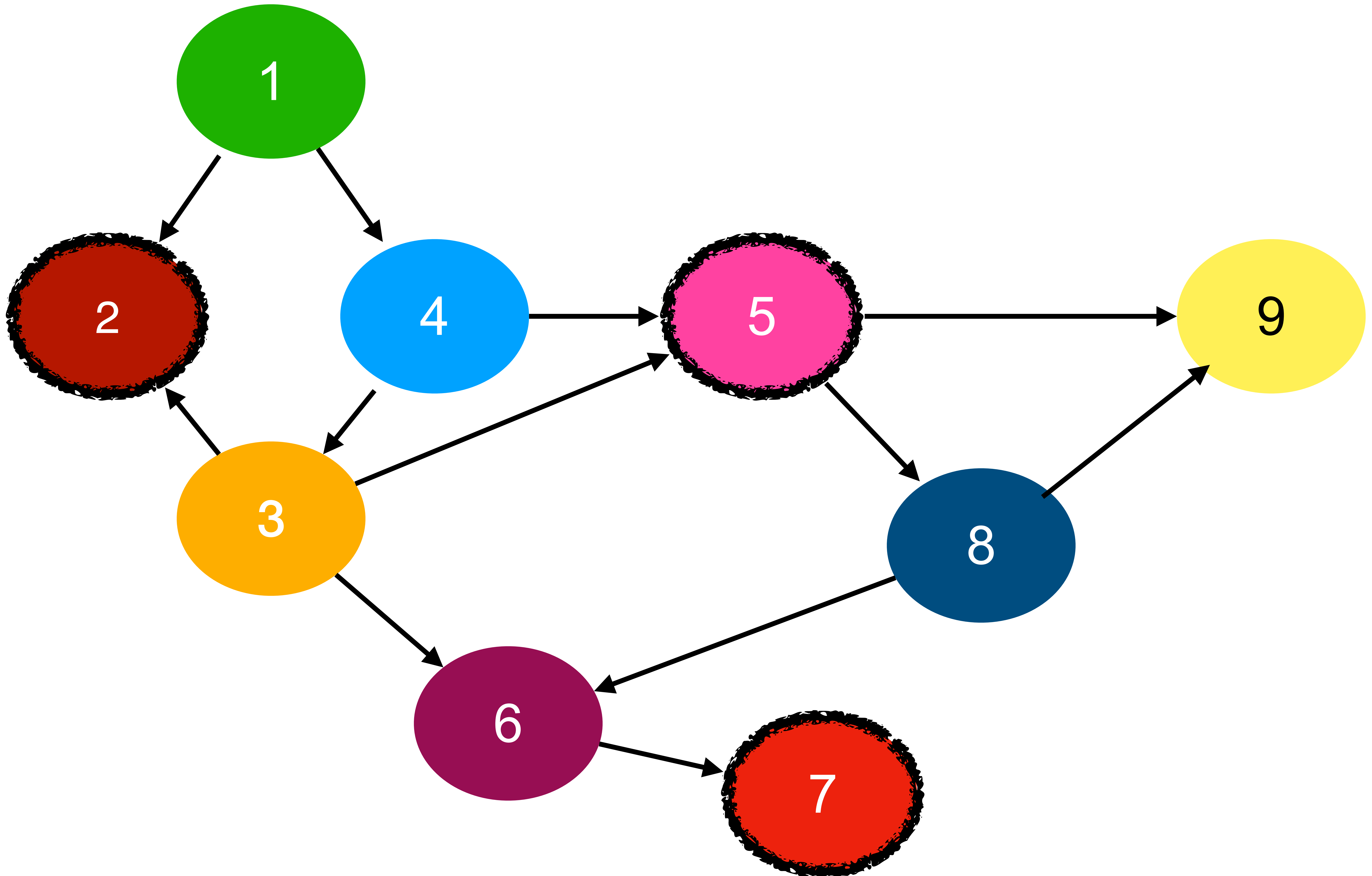
We will talk in detail how to learn the causal graph here:

1	Introduction
2	Probability recap
3	Graphical models, d-separation
4	Causal graphs, Interventions, SCMs
5	Covariate adjustment: backdoor criterion
6	Covariate Frontdoor criterion, Instrumental variables
7	Counterfactuals, potential outcomes, estimating causal effects 1
8	Estimating causal effects 2 (matching, IPW)
9	Constraint based structure learning
10	Score based structure learning, restricted models
11	Do-calculus, transportability, Joint Causal Inference
12	Causality-inspired ML, recap of the course

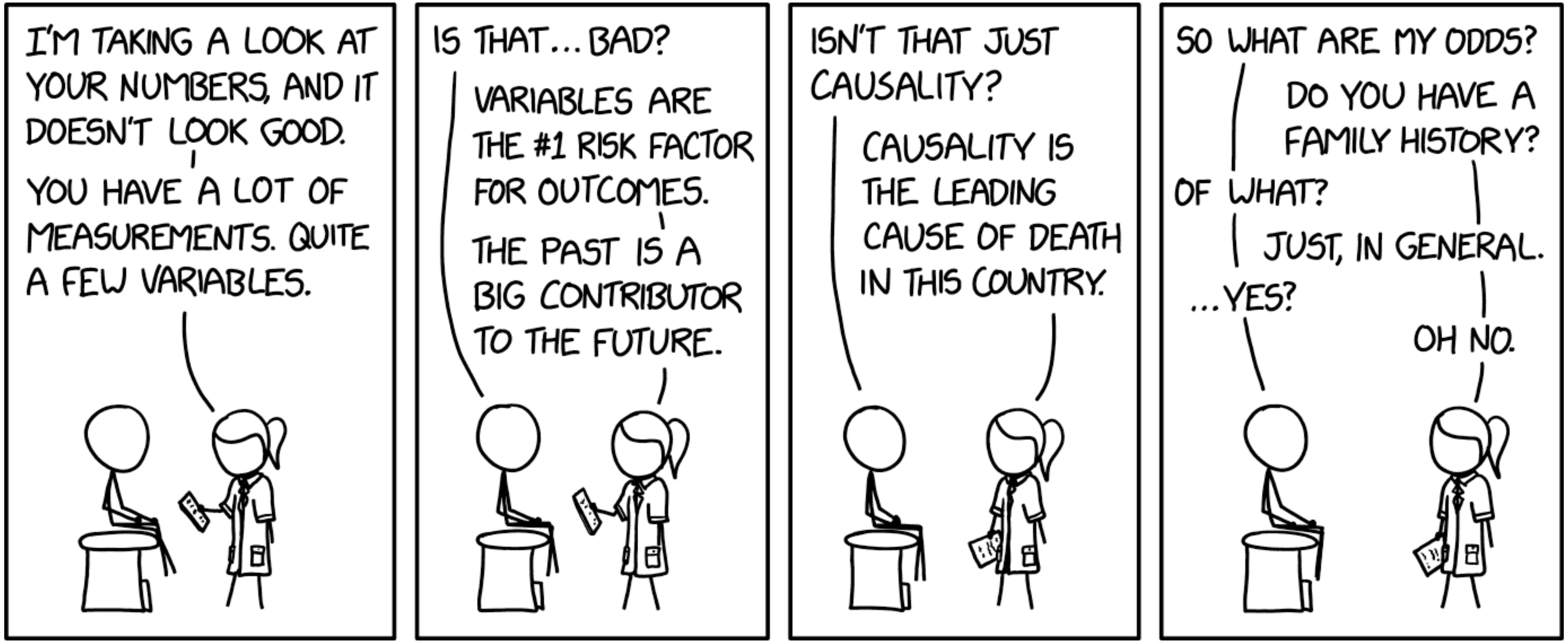
What happens if the graph is unknown?

For more details see: https://canvas.uva.nl/courses/36213/pages/course-schedule-tentative?module_item_id=1585152

Optional exercise in Canvas: d-separation 2



Questions??



https://xkcd.com/2620/