Chapter 3
Belief Networks

3.1 The Benefits of Structure

- Belief network is a way to depict the independence of variables

- 3.1.1 Modelling independencies

One morning Tracey leaves her house and realises that her grass is wet.
Is it due to overnight rain?
Or did she forget to turn off the sprinkler last night?

She notices that the grass of her neighbour, Jack, is also wet.
And she concludes that it has probably been raining.

$R \in \{0,1\}$ – rain
$S \in \{0,1\}$ – sprinkler
$J \in \{0,1\}$ – Jack's grass
$T \in \{0,1\}$ – Tracey's grass

$$
\begin{aligned}
p(T, J, R, S) &= p(T|J, R, S)\, p(J, R, S) \\
&= p(T|J, R, S)\, p(J|R, S)\, p(R, S) \\
&= p(T|J, R, S)\, p(J|R, S)\, p(R|S)\, p(S)
\end{aligned}
$$

$R \in \{0,1\}$ – rain
$S \in \{0,1\}$ – sprinkler
$J \in \{0,1\}$ – Jack's grass
$T \in \{0,1\}$ – Tracey's grass

$p(T|J, R, S)$ - $2^3 = 8$ values are needed
$p(J|R, S)$ - 4 values are needed
$p(R|S)$ - 2 values are needed
$p(S)$ - 1 value is needed
Total – 15 values are needs

- Conditional independence
  - Tracey's grass is wet
    - Depends on rain and her sprinkler.
    - Does not depend on Jack's grass

$$p(T|J,R,S) = p(T|R,S)$$

  - Jack's grass is wet
    - Depends on raining

$$p(J|R,S) = p(J|R)$$

  - The rain is not influenced by the sprinkler

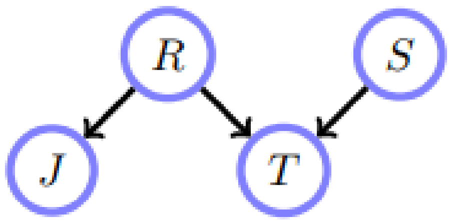
$$p(R|S) = p(R)$$

$$\begin{aligned} p(T,J,R,S) &= p(T|J,R,S)\,p(J,R,S) \\ &= p(T|J,R,S)\,p(J|R,S)\,p(R,S) \\ &= p(T|J,R,S)\,p(J|R,S)\,p(R|S)\,p(S \end{aligned}$$

$$p(T,J,R,S) = p(T|R,S)p(J|R)p(R)p(S)$$

$$p(T, J, R, S) = p(T|R, S)p(J|R)p(R)p(S)$$

$4 + 2 + 1 + 1 = 8$ values are needed



Belief network structure for the 'wet grass' example

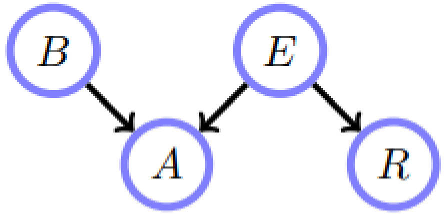
- conditional probability table(CPT)
  - $p(R = 1) = 0.2$
  - $p(S = 1) = 0.1$
  - $p(J = 1|R = 1) = 1$
    $p(J = 1|R = 0) = 0.2$
  - $p(T = 1|R = 1, S = 0) = 1$
    $p(T = 1|R = 1, S = 1) = 1$
    $p(T = 1|R = 0, S = 1) = 0.9$
    $p(T = 1|R = 0, S = 0) = 0$

$$
\begin{aligned}
p(S=1|T=1) &= \frac{p(S=1,T=1)}{p(T=1)} = \frac{\sum_{J,R} p(T=1,J,R,S=1)}{\sum_{J,R,S} p(T=1,J,R,S)} \\
&= \frac{\sum_{J,R} p(J|R)p(T=1|R,S=1)p(R)p(S=1)}{\sum_{J,R,S} p(J|R)p(T=1|R,S)p(R)p(S)} \\
&= \frac{\sum_{R} p(T=1|R,S=1)p(R)p(S=1)}{\sum_{R,S} p(T=1|R,S)p(R)p(S)} \\
&= \frac{0.9\times 0.8\times 0.1 + 1\times 0.2\times 0.1}{0.9\times 0.8\times 0.1 + 1\times 0.2\times 0.1 + 0\times 0.8\times 0.9 + 1\times 0.2\times 0.9} = 0.3382
\end{aligned}
$$

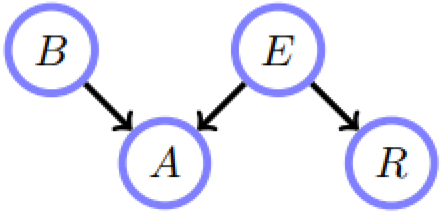
$$
\begin{aligned}
p(S=1|T=1,J=1) &= \frac{p(S=1,T=1,J=1)}{p(T=1,J=1)} \\
&= \frac{\sum_{R} p(T=1,J=1,R,S=1)}{\sum_{R,S} p(T=1,J=1,R,S)} \\
&= \frac{\sum_{R} p(J=1|R)p(T=1|R,S=1)p(R)p(S=1)}{\sum_{R,S} p(J=1|R)p(T=1|R,S)p(R)p(S)} \\
&= \frac{0.0344}{0.2144} = 0.1604
\end{aligned}
$$

- Example 3.1 (Was it the Burglar?)
  - The burglar alarm is sounding ($A = 1$)
  - Has she been burgled ($B = 1$),
    or
    was the alarm triggered by an earthquake ($E = 1$)?
  - The radio broadcasts an earthquake alert ($R = 1$).

Was it the Burglar?
(1) $p(B = 1|A = 1) = ?$
(2) $p(B = 1|A = 1, R = 1) = ?$



$$
\begin{aligned}
p(B,E,A,R) &= p(A|B,E,R)\,p(B,E,R) \\
&= p(A|B,E,R)\,p(R|B,E)\,p(E|B)\,p(B) \\
&= p(A|B,E)\,p(R|E)\,p(E)\,p(B)
\end{aligned}
$$



Specifying conditional probability tables

Alarm = 1	Burglar	Earthquake
0.9999	1	1
0.99	1	0
0.99	0	1
0.0001	0	0

Radio = 1	Earthquake
1	1
0	0

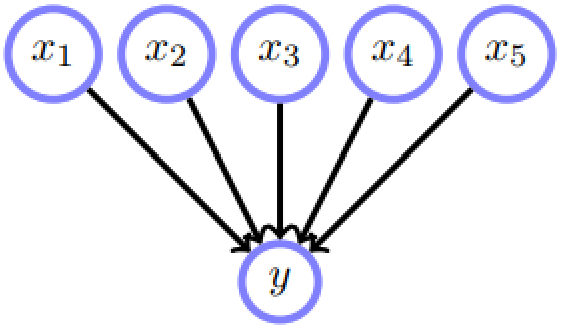
$p(B=1)=0.01\,,\ p(E=1)=0.000001$

$$
\begin{aligned}
p(B=1|A=1) &= \frac{\sum_{E,R} p(B=1,E,A=1,R)}{\sum_{B,E,R} p(B,E,A=1,R)} \\
&= \frac{\sum_{E,R} p(A=1|B=1,E)p(B=1)p(E)p(R|E)}{\sum_{B,E,R} p(A=1|B,E)p(B)p(E)p(R|E)} \approx 0.99
\end{aligned}
$$

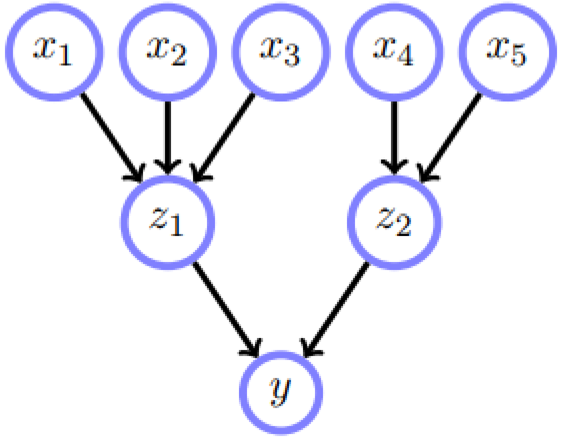
$p(B=1|A=1,R=1) \approx 0.01$

- 3.1.2 Reducing the burden of specification

Assume that all variables are binary



$2^5 = 32$ states



$2^3 + 2^2 + 2^2 = 16$ states

$$p(y|x_1, \ldots, x_5) = \sum_{z_1, z_2} p(y|z_1, z_2) p(z_1|x_1, x_2, x_3) p(z_2|x_4, x_5)$$

3.2 Uncertain and Unreliable Evidence

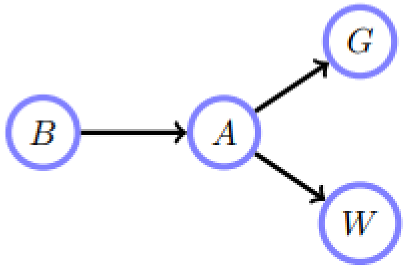
- 3.2.1 Uncertain evidence
  - Soft or uncertain evidence
    - The strength of our belief
    - Example) Sally heard the burglar alarm. But we are 70% sure that the burglar alarm went off.
  - Hard evidence
    - We are certain that a variable is in a particular state.
    - 100% or 0%

- Example 3.2 (soft-evidence)
  - We are only 70% sure we heard the burglar alarm sounding
  - Soft-evidence variable $\tilde{A} = (0.7, 0.3)$
  - Meaning

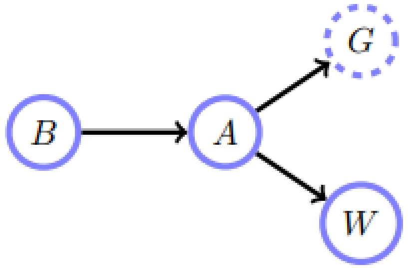
$$p\left(A = 1 \middle| \tilde{A}\right) = 0.7 \text{ and } p\left(A = 0 \middle| \tilde{A}\right) = 0.3$$

  - Target: $p\left(B = 1 \middle| \tilde{A}\right) \approx 0.6930$

- Holmes, Watson and Mrs Gibbon
  - Watson and Mrs Gibbon are neighbors of Holms
  - Watson heard the alarm is sounding
  - Mrs Gibbon heard the alarm
    - She is a little deaf and cannot be sure herself
    - 80% sure
  - Was Holmes' house burgled?



hard evidence



soft evidence

- Holmes, Watson and Mrs Gibbon

$B \in \{\text{tr}, \text{fa}\}$ $B = \text{tr}$ means that Holmes' house has been burgled
$A \in \{\text{tr}, \text{fa}\}$ $A = \text{tr}$ means that Holmes' house Alarm went off
$W \in \{\text{tr}, \text{fa}\}$ $W = \text{tr}$ means that Watson heard the alarm
$G \in \{\text{tr}, \text{fa}\}$ $G = \text{tr}$ means that Mrs Gibbon heard the alarm

hard evidence

$$p(B, A, G, W) = p(A|B)p(B)p(W|A)p(G|A)$$

$$p(B = \text{tr}|W = \text{tr}, G) = \frac{p(B = \text{tr}, W = \text{tr}, G)}{p(W = \text{tr}, G)} = \frac{\sum_A p(G|A)p(W = \text{tr}|A)p(A|B = \text{tr})p(B = \text{tr})}{\sum_{B,A} p(G|A)p(W = \text{tr}|A)p(A|B)p(B)}$$

soft evidence

$$p\left(B = \text{tr}\middle|W = \text{tr}, \tilde{G}\right) = p(B = \text{tr}|W = \text{tr}, G = \text{tr})\, p\left(G = \text{tr}\middle|\tilde{G}\right) + p(B = \text{tr}|W = \text{tr}, G = \text{fa})\, p(G = \text{fa}|\tilde{G})$$

$$p\left(G = \text{tr}\middle|\tilde{G}\right) = 0.8,\ p\left(G = \text{fa}\middle|\tilde{G}\right) = 0.2$$

- 3.2.2 Unreliable evidence

3.3 Belief Networks

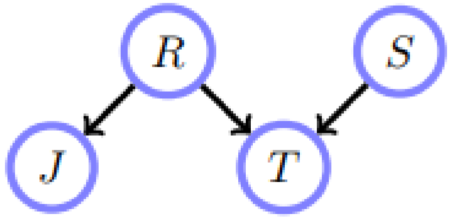
- Definition 3.1 (Belief network)
  - A belief network is a distribution of the form

$$p(x_1, \ldots, x_D) = \prod_{i=1}^{D} p(x_i | \mathrm{pa}(x_i))$$

    where $\mathrm{pa}(x_i)$ represents the parental variables of x_i
  - Belief network corresponds to a Directed Acyclic Graph(DAG)

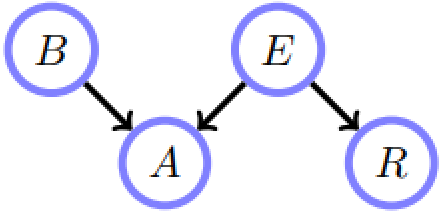
- Wet grass

$$P(J, R, T, S) = p(J|R)\, p(R)\, p(T|R, S)\, p(S)$$

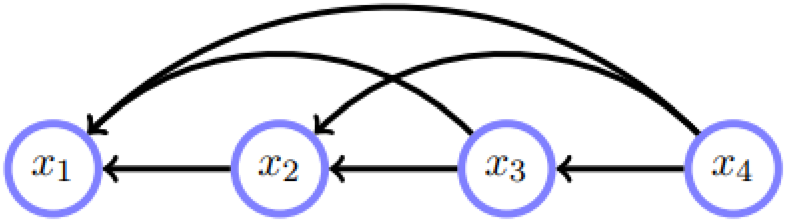


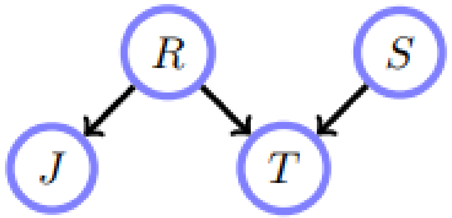
- Burglar

$$P(B, A, E, R) = p(B)\, p(A|B, E)\, p(E)\, p(R|E)$$



- DAG may be written in the cascade form

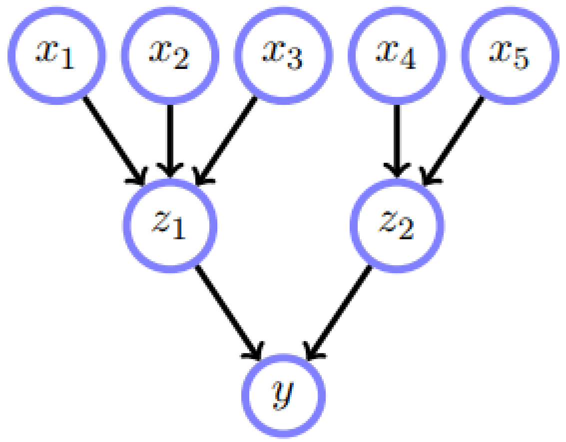




$x_1 = J,\ x_2 = T,\ x_3 = R,\ x_4 = S$

$$
\begin{aligned}
p(x_1, \ldots, x_n) &= p(x_1 | x_2, \ldots, x_n) p(x_2, \ldots, x_n) \\
&= p(x_1 | x_2, \ldots, x_n) p(x_2 | x_3, \ldots, x_n) p(x_3, \ldots, x_n) \\
&= p(x_n) \prod_{i=1}^{n-1} p(x_i | x_{i+1}, \ldots, x_n)
\end{aligned}
$$

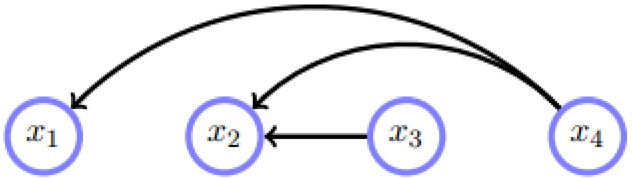
Remark 3.3 (Dependencies and the Markov Blanket)**.** Consider a distribution on a set of variables $\mathcal{X}$. For a variable $x_i \in \mathcal{X}$ and corresponding belief network represented by a DAG G, let $MB(x_i)$ be the variables in the Markov blanket of x_i. Then for any other variable y that is also not in the Markov blanket of x_i ($y \in \mathcal{X} \backslash \{x_i \cup MB(x_i)\}$), then $x_i \perp\!\!\!\perp y | MB(x_i)$. That is, the Markov blanket of x_i carries all information about x_i. As an example, for fig(3.2b), $MB(z_1) = \{x_1, x_2, x_3, y, z_2\}$ and $z_1 \perp\!\!\!\perp x_4 | MB(z_1)$.



$$p(z_1, x_4 | x_1, x_2, x_3, y, z_2) = p(z_1 | x_1, x_2, x_3, y, z_2)\ p(x_4 | x_1, x_2, x_3, y, z_2)\ ?$$

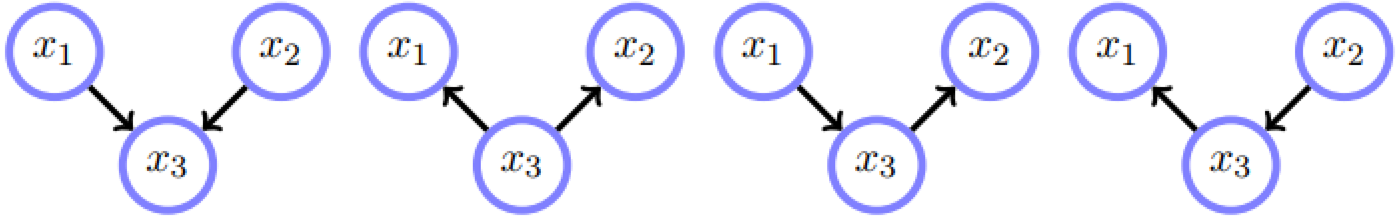
- 3.3.1 Conditional independence

Whilst a BN corresponds to a set of conditional independence assumptions, it is not always immediately clear from the DAG whether a set of variables is conditionally independent of a set of other variables



Are x_1 and x_2 independent, given the state of x_4?

$p(x_1, x_2|x_4) = p(x_1|x_4)p(x_2|x_4)$?

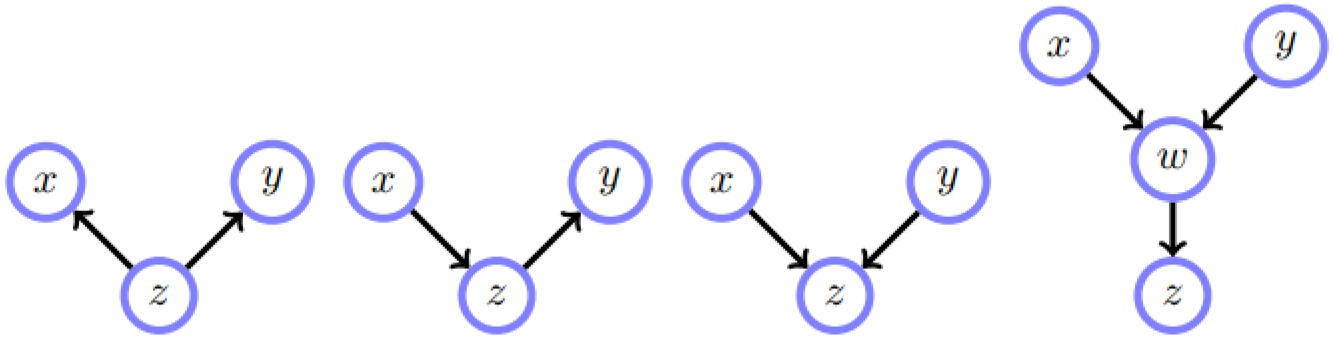


(a) (b) (c) (d)

Are any of these graphs equivalent, in the sense that they represent the same distribution?
Answer: $x_1 \coprod x_2 | x_3$

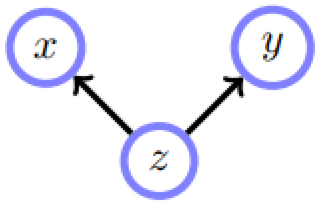
- 3.3.2 The impact of collisions

- Definition 3.2. Given a path $\mathcal{P}$, a collider is a node c on $\mathcal{P}$ with neighbours a and b on $\mathcal{P}$ such that $a \rightarrow c \leftarrow b$.



(a) (b) (c) (d)

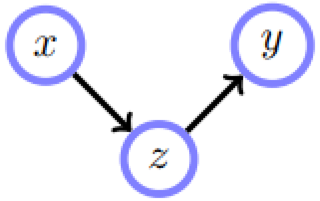
- How can we check if $x \coprod y \mid z$?



(a)

$$p(x, y|z) = \frac{p(x,y,z)}{p(z)}$$
$$= \frac{p(x|z)p(y|z)\,p(z)}{p(z)}$$
$$= p(x|z)p(y|z)$$

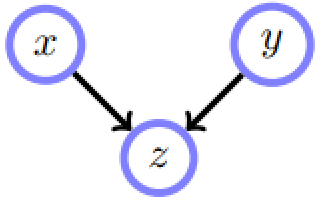
Independent conditioned on z



(b)

$$p(x, y|z) = \frac{p(x,y,z)}{p(z)}$$
$$= \frac{p(y|z)p(z|x)\,p(x)}{p(z)}$$
$$= p(y|z)\frac{p(z|x)\,p(x)}{p(z)}$$
$$= p(y|z)p(x|z)$$

Independent conditioned on z

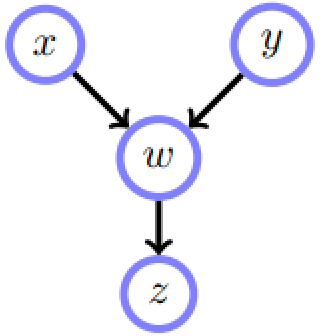


(c)

$$p(x,y|z) = \frac{p(x,y,z)}{p(z)} = \frac{p(z|x,y)p(x)p(y)}{p(z)}$$

$$p(x|z)p(y|z) = \frac{p(x,z)}{p(z)}\frac{p(y,z)}{p(z)} = \frac{\sum_{y'} p(x,y',z)}{p(z)}\frac{\sum_{x'} p(x',y,z)}{p(z)} = \frac{\sum_{y'} p(z|x,y')p(x)p(y')}{p(z)}\frac{\sum_{x'} p(z|x',y)p(x')p(y)}{p(z)}$$

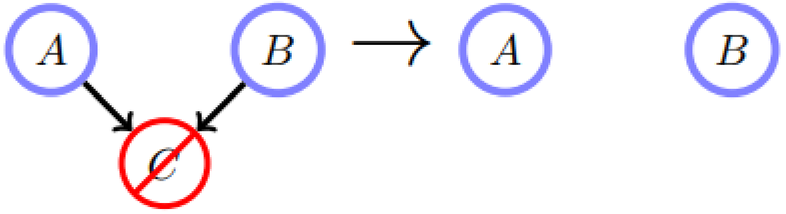
Graphically dependent conditioned on z
i.e. may not be independent



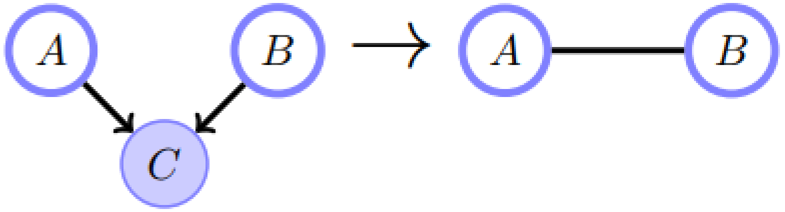
(d)

$$p(x,y|z) = \frac{p(x,y,z)}{p(z)} = \frac{\sum_w p(x,y,z,w)}{p(z)} = \frac{\sum_w p(z|w)p(w|x,y)p(x)p(y)}{p(z)} \neq p(x|z)p(y|z)$$

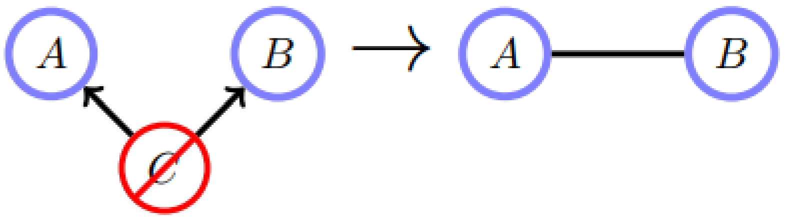
Graphically dependent conditioned on z
i.e. may not be independent



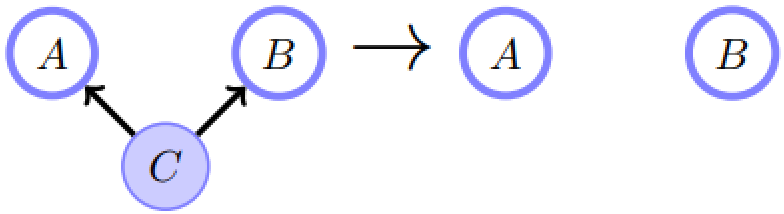
Marginalising over C makes A and B independent.
A and B are independent: $p(A, B) = p(A)p(B)$



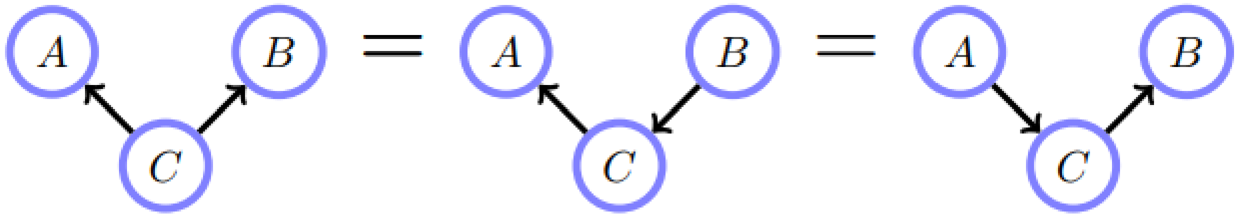
Conditioning on C makes A and B (graphically) dependent, in general $p(A, B|C) \neq p(A|C)\, p(B|C)$



Marginalising over C makes A and B dependent.
in general $p(A,B) \neq p(A)p(B)$



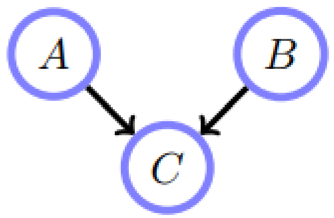
Conditioning on C makes A and B independent:
$p(A,B|C) = p(A|C)p(B|C)$



These graphs all express the same
conditional independence assumptions

- 3.3.3 Graphical path manipulations for independence

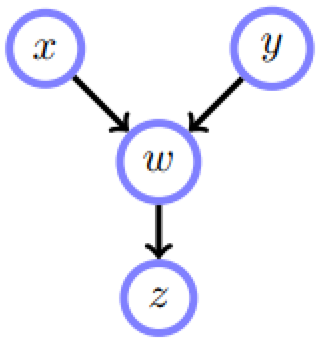
- 3.3.4 d-Separation



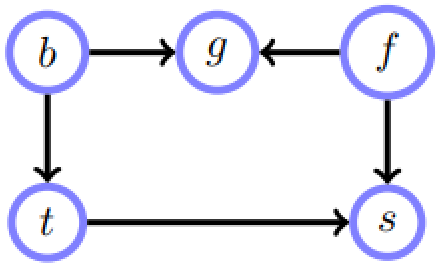
$\mathcal{X} = \{A\}$
$\mathcal{Y} = \{B\}$
$\mathcal{Z} = \{C\}$

- Definition 3.4 (d-connection, d-separation)
  - G is a directed graph
  - $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are disjoint sets of vertices
  - $\mathcal{X}$ and $\mathcal{Y}$ are said to be d-connected by $\mathcal{Z}$ if
    - There exists an undirected path U
    - between some vertex in $\mathcal{X}$ and some vertex in $\mathcal{Y}$
    - for every collider C on U, either C or a descendent of C is in $\mathcal{Z}$
    - no non-collider on U is in $\mathcal{Z}$
  - $\mathcal{X}$ and $\mathcal{Y}$ are said to be d-separated by $\mathcal{Z}$ if
    - they are not d-connected

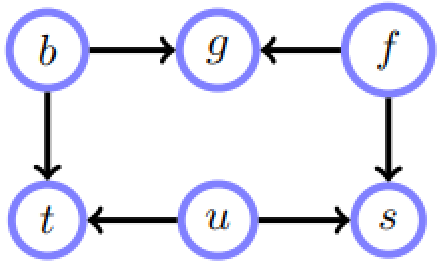
- $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z} \Rightarrow \mathcal{X} \perp\!\!\!\perp \mathcal{Y} \mid \mathcal{Z}$



$\mathcal{X} = \{A\}$, $\mathcal{Y} = \{B\}, \mathcal{Z} = \{C\}$
Are $\mathcal{X}$ and $\mathcal{Y}$ d-connected by $\mathcal{Z}$?



t and f are
d-connected by g



b and f are
d-separated by u

- 3.3.5 Graphical and distributional in/dependence
  - Graphical independence ⇒ distributional independence
  - Graphical dependence ⇒ distributional dependence ?
    - Example

      $a \to c \leftarrow b \;\Rightarrow\; a \coprod b$, $a \prod c$ and $b \prod c$ graphically, not distributionally

      $\mathrm{dom}(a) = \mathrm{dom}(b) = \mathrm{dom}(c) = \{0,1\}$

      (1) $p_1(c = 1|a, b) = (a - b)^2$, $p_1(a = 1) = 0.3$, $p_1(b = 1) = 0.4$

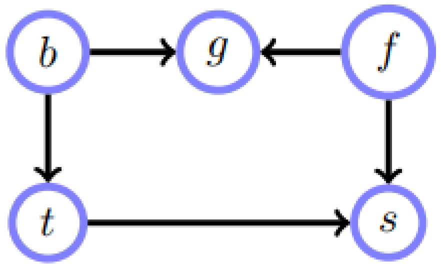
      $\Rightarrow\; a \coprod b$, $a \prod c$, $b \prod c$

      (2) $p_2(c = 1|a, b) = 0.5$, $p_2(a = 1) = 0.3$, $p_2(b = 1) = 0.4$

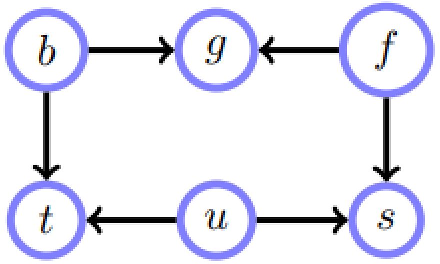
      $\Rightarrow\; a \coprod b$, $a \coprod c$, $b \coprod c$

- Example 3.3
  - $t \coprod f | \emptyset$?
    t and f are d-separated (by $\emptyset$)

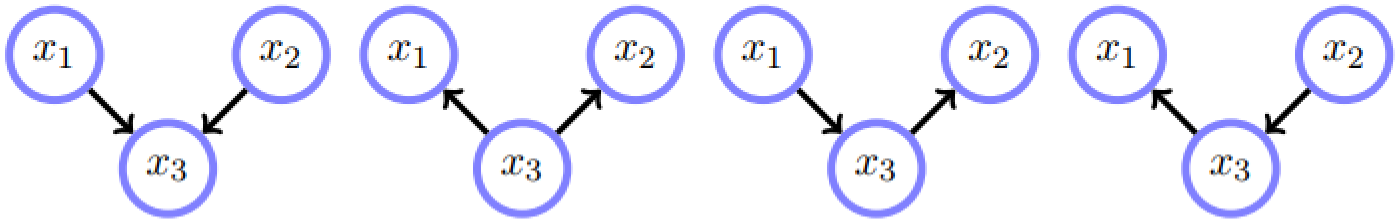
  - $t \coprod f | g$?
    t and f are d-connected by g



- Example 3.4
  - $\{b, f\} \coprod u | \emptyset$?
    $\{b, f\}$ and u are d-separated (by $\emptyset$)



- 3.3.6 Markov equivalence in belief networks



(a) (b) (c) (d)

Possible independence statements

$$x_1 \coprod x_2,\ x_1 \coprod x_2 | x_3$$
$$x_1 \coprod x_3,\ x_1 \coprod x_3 | x_2$$
$$x_2 \coprod x_3,\ x_2 \coprod x_3 | x_2$$
$$\{x_1, x_2\} \coprod x_3$$
$$\{x_1, x_3\} \coprod x_2$$
$$\{x_2, x_3\} \coprod x_1$$

Definition 3.5 (Markov equivalence).
Two graphs are Markov equivalent if they represent the same set of conditional independence statements.

- Procedure 3.1 (Markov equivalence)
  - Three nodes, A, B, C are immoral if
    - C is a child of both A and B
    - A and B not directly connected
  - skeleton of a graph
    - removing the directions on the arrows
  - Two DAGs represent the same set of independence assumptions if and only if they have the same skeleton and the same set of immoralities

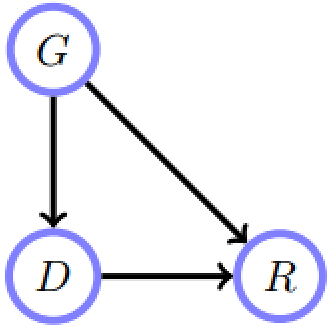
3.4 Causality

- 3.4.1 Simpson's paradox

Males	Recovered	Not Recovered	Rec. Rate
Given Drug	18	12	60%
Not Given Drug	7	3	70%

Females	Recovered	Not Recovered	Rec. Rate
Given Drug	2	8	20%
Not Given Drug	9	21	30%

Combined	Recovered	Not Recovered	Rec. Rate
Given Drug	20	20	50%
Not Given Drug	16	24	40%



A model of Gender, Drug and Recovery

$$p(G,D,R) = p(R|G,D)p(D|G)p(G)$$

$$p(R|D = \mathrm{tr}) = 0.5$$

$$p(R|D = \mathrm{fa}) = 0.4$$

Modified model

$$p(G,D,R) = p(R|G,D)p(D|G)p(G)$$

$$p(D|G) = 0.5$$

Modified model

$$p(G,D,R) = p(R|G,D)p(D|G)p(G)$$
$$p(D|G) = 0.5$$

Males	Recovered	Not Recovered	Rec. Rate
Given Drug	18	12	60%
Not Given Drug	7	3	70%

Females	Recovered	Not Recovered	Rec. Rate
Given Drug	2	8	20%
Not Given Drug	9	21	30%

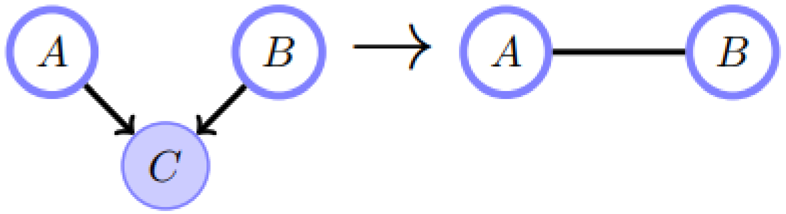
Combined	Recovered	Not Recovered	Rec. Rate
Given Drug	20	20	50%
Not Given Drug	16	24	40%

$$P(R = \text{tr}|D = \text{tr}) = \frac{p(D{=}\text{tr},R{=}\text{tr})}{p(D{=}\text{tr})}$$
$$= \frac{\sum_G p(G,R{=}\text{tr},D{=}\text{tr})}{\sum_{G,R} p(G,R,D{=}\text{tr})}$$
$$= \frac{\sum_G p(R{=}\text{tr}|G,D{=}\text{tr})p(D{=}\text{tr}|G)p(G)}{\sum_{G,R} p(R|G,D{=}\text{tr})p(D{=}\text{tr}|G)p(G)}$$
$$= \frac{p(R{=}\text{tr}\,|G{=}\text{ma},D{=}\text{tr})\,p(D{=}\text{tr}|G{=}\text{ma})\,p(G{=}\text{ma}) + p(R{=}\text{tr}|G{=}\text{fe},D{=}\text{tr})\,p(D{=}\text{tr}|G{=}\text{fe})\,p(G{=}\text{fe})}{\sum_G p(D{=}\text{tr}|G)\,p(G)}$$
$$= \frac{0.6 \times 0.5 \times 0.5 + 0.2 \times 0.5 \times 0.5}{0.5 \times 0.5 + 0.5 \times 0.5}$$
$$= 0.4$$

$$P(R = \text{tr}|D = \text{fa}) = 0.5$$

Summary

- Belief network $$p(x_1, \dots, x_D) = \prod_{i=1}^{D} p(x_i | \mathrm{pa}(x_i))$$
- Collider



$$p(A, B | C) \neq p(A|C)\, p(B|C)$$

- $\mathcal{X}$ and $\mathcal{Y}$ are d-separated by $\mathcal{Z}$ $\Rightarrow$ $\mathcal{X} \coprod \mathcal{Y} \,|\, \mathcal{Z}$
- Markov equivalence: Two DAGs represent the same set of independence assumptions if and only if they have the same skeleton and the same set of immoralities

3.7 Exercises

Exercise 3.1 (Party Animal). *The party animal problem corresponds to the network in fig(3.14). The boss is angry and the worker has a headache – what is the probability the worker has been to a party? To complete the specifications, the probabilities are given as follows:*

$$p(U = \mathsf{tr} | P = \mathsf{tr}, D = \mathsf{tr}) = 0.999 \quad p(U = \mathsf{tr} | P = \mathsf{fa}, D = \mathsf{tr}) = 0.9 \quad p(H = \mathsf{tr} | P = \mathsf{tr}) = 0.9$$
$$p(U = \mathsf{tr} | P = \mathsf{tr}, D = \mathsf{fa}) = 0.9 \quad p(U = \mathsf{tr} | P = \mathsf{fa}, D = \mathsf{fa}) = 0.01 \quad p(H = \mathsf{tr} | P = \mathsf{fa}) = 0.2$$
$$p(A = \mathsf{tr} | U = \mathsf{tr}) = 0.95 \quad p(A = \mathsf{tr} | U = \mathsf{fa}) = 0.5 \quad p(P = \mathsf{tr}) = 0.2, p(D = \mathsf{tr}) = 0.4$$

$$p(P = \mathrm{tr} \mid A = \mathrm{tr}, H = \mathrm{tr}) = ?$$

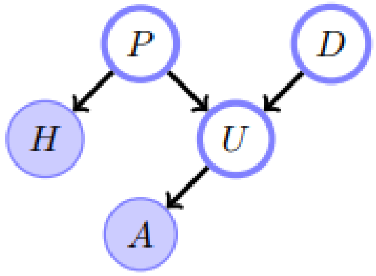


Figure 3.14: Party animal. Here all variables are binary. P = Been to Party, H = Got a Headache, D = Demotivated at work, U = Underperform at work, A =Boss Angry. Shaded variables are observed in the true state.

- Exercise 3.3

State if the following conditional independence relationships are true or false

1. $t \amalg s \mid d$
2. $l \amalg b \mid s$
3. $a \amalg s \mid l$
4. $a \amalg s \mid \{l, d\}$