BAYESIAN NETWORKS 1

ARTIFICIAL INTELLIGENCE | COMP 131

TODAY ON AI

- Bayesian networks
- Independence
- Exact inference
- Questions?

- **Conditional probability**: $P(X \mid Y) = \frac{P(X,Y)}{P(Y)}$

- **Product rule**: $P(X,Y) = P(X \mid Y)\, P(Y) = P(Y \mid X)\, P(X)$

- **Chain rule**: $P(X_1, \dots, X_n) = \prod_i P(X_i \mid X_1, \dots, X_{i-1})$

- **X and Y are independent**: iff $P(X,Y) = P(X)\, P(Y)$

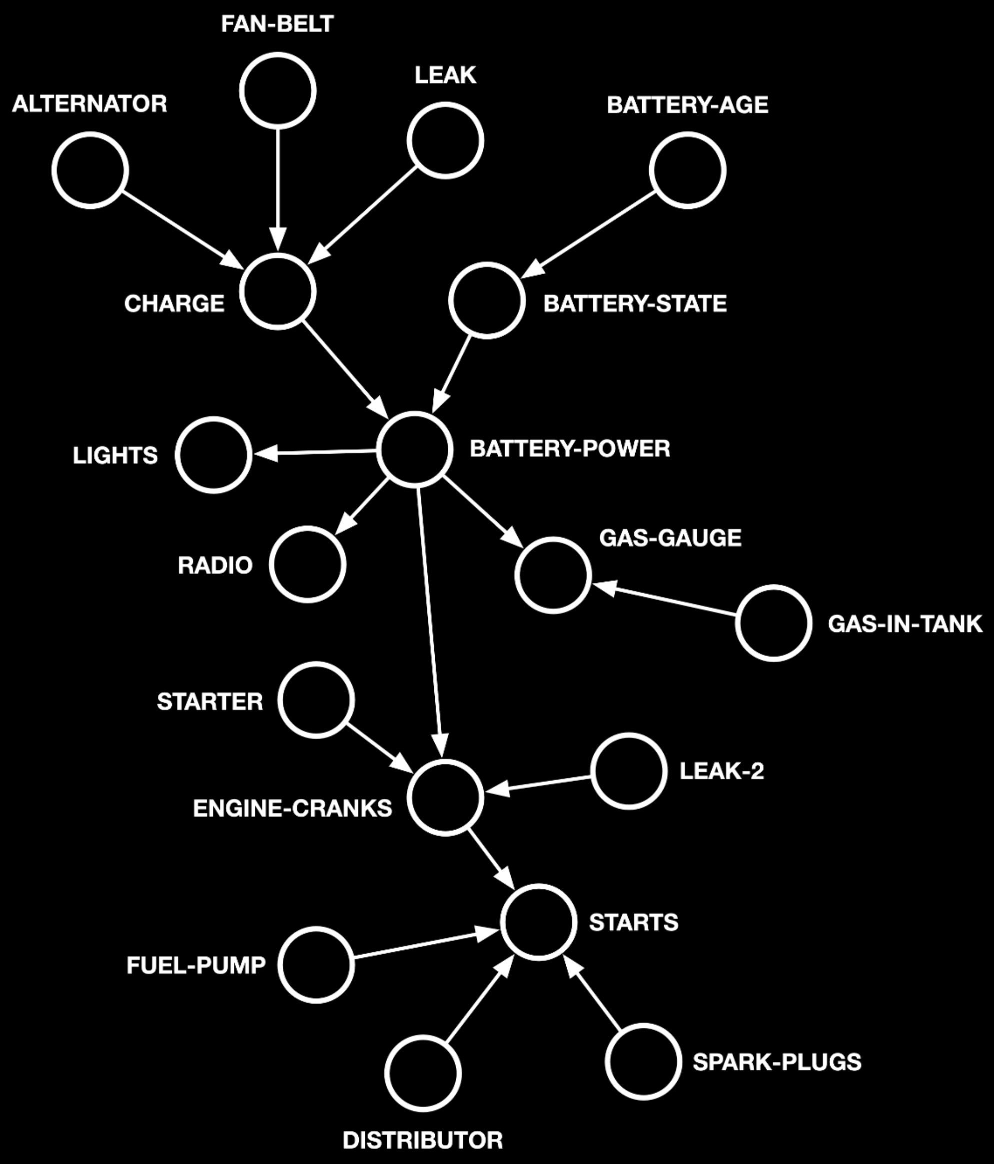
- **Bayes rule**: $P(Cause \mid Observation) = \frac{P(Observation \mid Cause) P(Cause)}{P(Observation)}$

SECTION 01

Bayesian networks

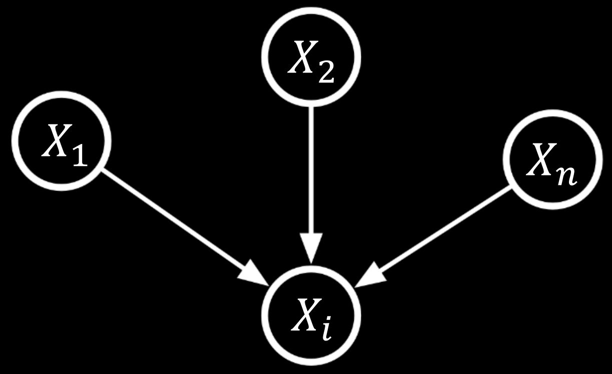
Bayesian networks, or **Belief networks** or more formally **graphical models**, are a simplified descriptions of how some portion of the world work:

- It is a compact way to describe joint probabilities
- It allows to calculate complex joint distributions using **local conditional probabilities** among random variables
- Local interactions will **chain together** to give global, indirect interactions



- **Node**: It represents a **variable** with its domain:
  - Can be assigned (observed) or unassigned (unobserved)
  - There is usually one node per random variable

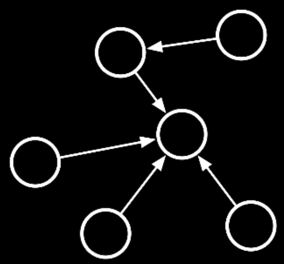
- **Arc**: It encodes an **interaction** or local conditional probability between variables



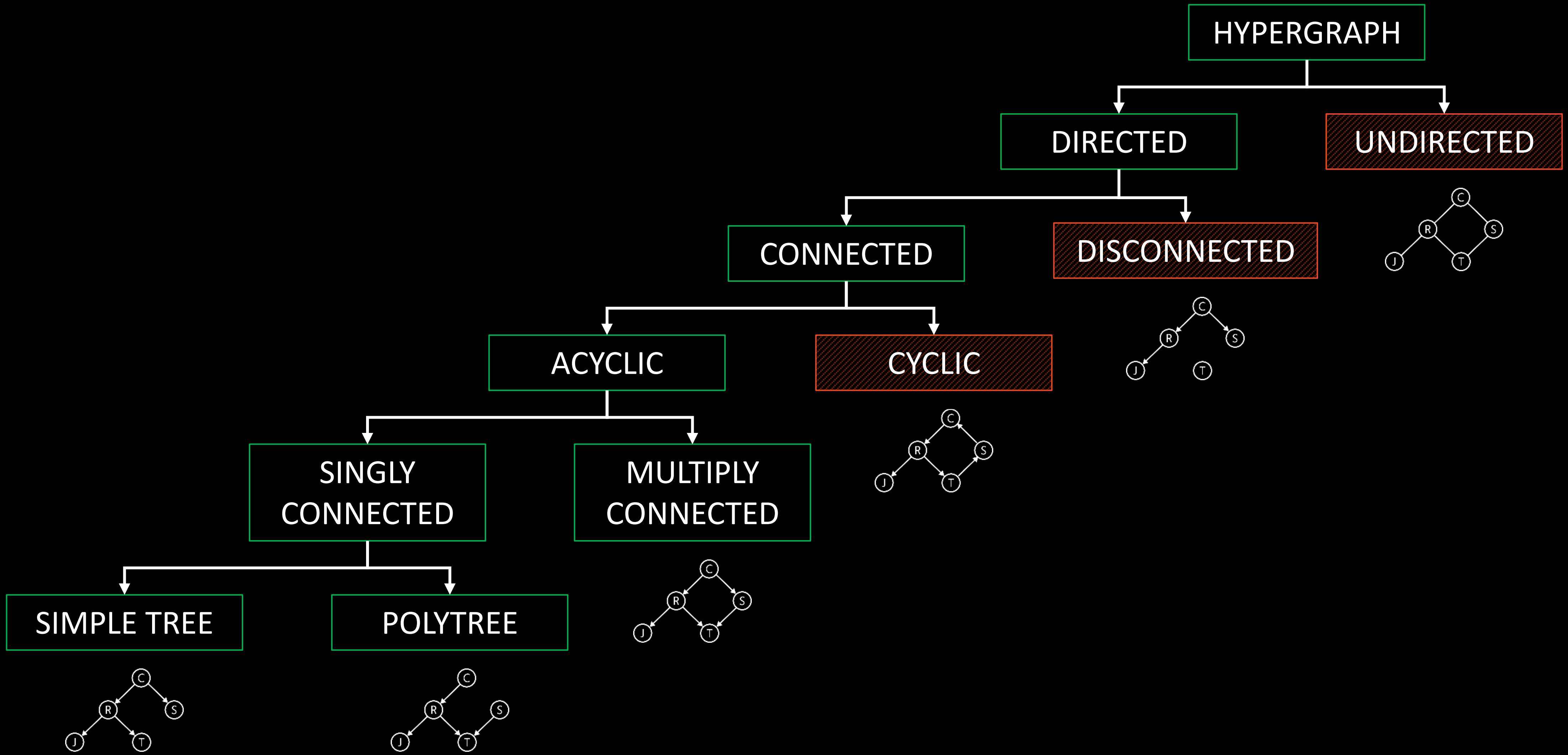
$$P(X_i \mid X_1, \dots, X_n) = P(x_i \mid parents(X_i))$$

- **Nodes without arcs**: They represent **independent** random variables

- **Hypergraph**: A directed, acyclic graph that encodes **conditional independence**, and globally describe a **joint probability**

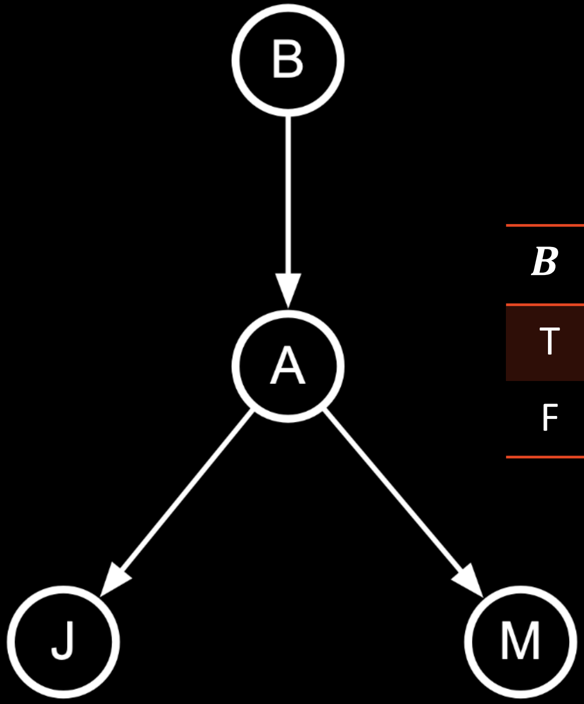


$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$


HYPERGRAPH
DIRECTED
UNDIRECTED
CONNECTED
DISCONNECTED
ACYCLIC
CYCLIC
SINGLY CONNECTED
MULTIPLY CONNECTED
SIMPLE TREE
POLYTREE

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999



$$P(B, A, J, M) = P(B)\, P(A|B)\, P(J|A)\, P(M|A)$$

J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

B	$P(a\|B)$	$P(\neg a\|B)$
T	0.95	0.05
F	0.29	0.71

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

A useful and fundamental condition that Bayesian networks capture is called **conditional independence** and indicated as $\boldsymbol{X} \perp \boldsymbol{Y} \mid \boldsymbol{Z}$:

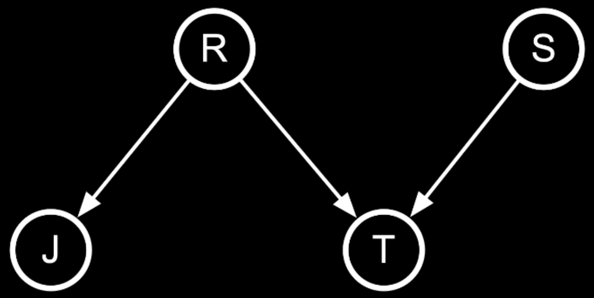
$$\forall x \in X, y \in Y, \forall z \in Z: P(x, y \mid z) = P(x \mid z)\, P(y \mid z)$$

$$\forall x \in X, y \in Y, z \in Z: P(x \mid z, y) = P(x \mid z)$$

$$\forall x \in X, y \in Y, z \in Z: P(y \mid z, x) = P(y \mid z)$$

One morning Tracey leaves her house and realizes that her grass is wet. Is it due to overnight rain or did she forget to turn off the sprinkler last night?
Next, she notices that the grass of her neighbor, Jack, is also wet.

Example:
$R = Rained$
$S = Sprinkler$
$J = Jack's\ grass\ is\ wet$
$T = Tracey's\ grass\ is\ wet$



A **good conditional independence** assumption is: $J \perp T \mid R$ that is:

$$P(J, T \mid \mathrm{R}) = P(J \mid R)\, P(T \mid R)$$

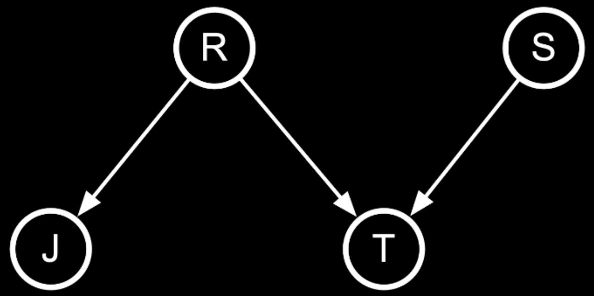
The chain rule says: $P(x_1, \dots, x_n) = \prod_i P(x_i | x_1, \dots, x_{i-1})$

Let's look at the previous example:

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

If we assume that $J \perp T \mid R$ the conditional probabilities are simplified:

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



$$P(J \mid R,T) = P(J \mid R)\ ?$$

$$P(J \mid R,T) = \frac{P(J,T,R)}{P(R,T)} \quad P(J,T,R) = P(J,T \mid R)\, P(R) \quad P(J,T \mid R) = P(J \mid R)\, P(T \mid R) \quad P(R,T) = P(T \mid R)\, P(R)$$

$$P(J \mid R,T) = \frac{P(J \mid R)\, P(T \mid R)\, P(R)}{P(T \mid R)\, P(R)} \qquad \boxed{P(J \mid R,T) = P(J \mid R)}$$

$$P(T \mid J,R,S) = P(T \mid R,S)\ ?$$

$$P(T \mid J,R,S) = \frac{P(J,S,T,R)}{P(J,R,S)} \quad P(J,S,T,R) = P(J,T \mid R,S)\, P(R,S) \quad P(J,T \mid R,S) = P(J \mid R,S)\, P(T \mid R,S)$$

$$P(J,R,S) = P(J \mid R,S)\, P(R,S)$$

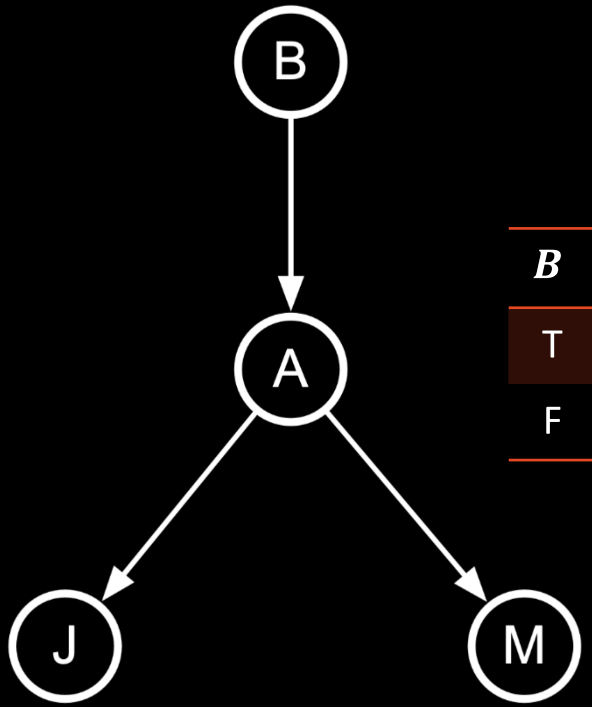
$$P(T \mid J,R,S) = \frac{P(J \mid R,S)\, P(T \mid R,S)\, P(R,S)}{P(J \mid R,S)\, P(R,S)} \qquad \boxed{P(T \mid J,R,S) = P(T \mid R,S)}$$

SECTION 02

Exact inference

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a\|B)$	$P(\neg a\|B)$
T	0.95	0.05
F	0.29	0.71

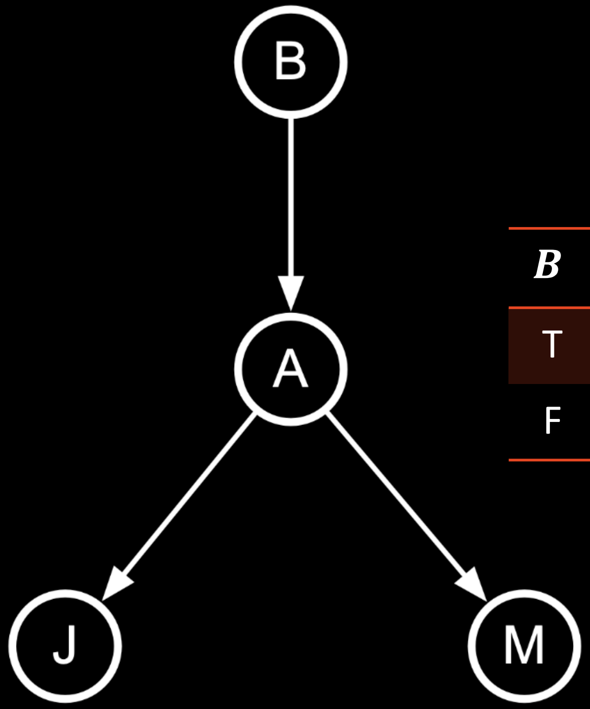
J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a\|B)$	$P(\neg a\|B)$
T	0.95	0.05
F	0.29	0.71

$$P(B, A, J, M) = P(B)\, P(A|B)\, P(J|A)\, P(M|A)$$

$$P(b, a, \neg j, m) = ?$$

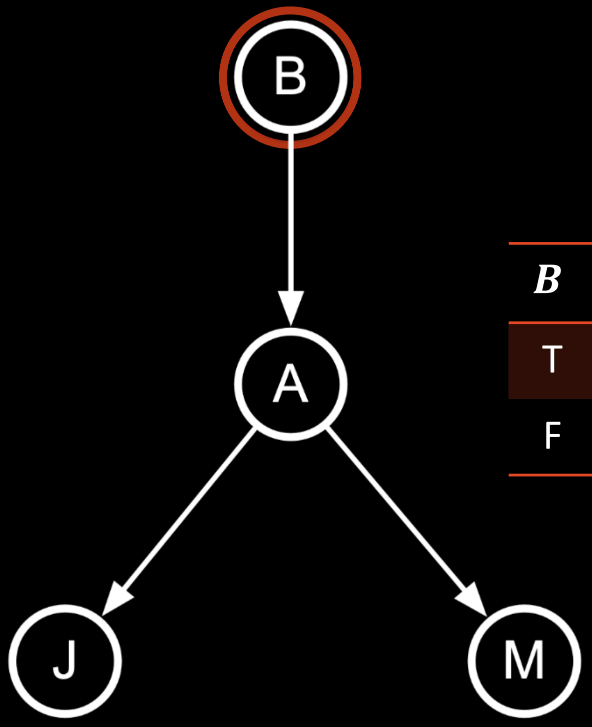
J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a \mid B)$	$P(\neg a \mid B)$
T	0.95	0.05
F	0.29	0.71

$$P(B, A, J, M) = P(B)\, P(A|B)\, P(J|A)\, P(M|A)$$

$$P(b, a, \neg j, m) = P(b) \;\; \times P(a|b) \times P(\neg j|a) \times P(m|a)$$
$$P(b, a, \neg j, m) = \mathbf{0.001}$$

$$P(b, a, \neg j, m) = \mathbf{0.001}$$

J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

A	$P(j \mid A)$	$P(\neg j \mid A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m \mid A)$	$P(\neg m \mid A)$
T	0.70	0.30
F	0.01	0.99

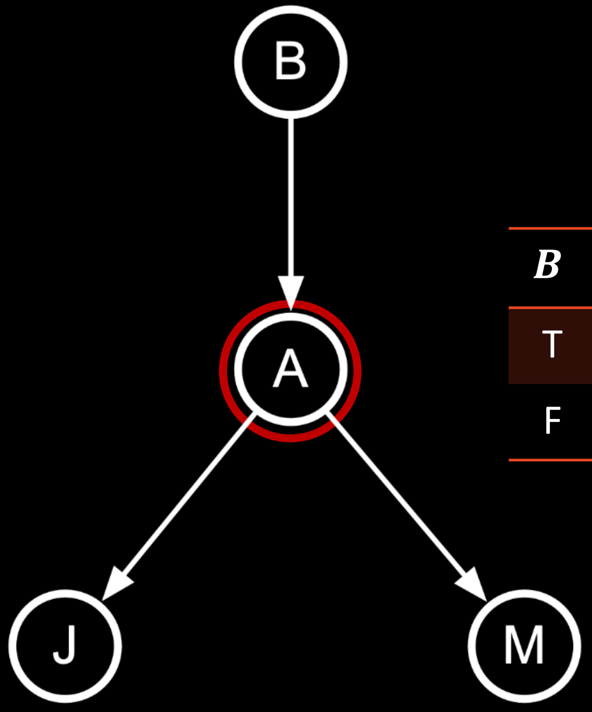
You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$$P(B, A, J, M) = P(B)\, P(A|B)\, P(J|A)\, P(M|A)$$

$$P(b, a, \neg j, m) = P(b) \times P(a|b) \times P(\neg j|a) \times P(m|a)$$
$$P(b, a, \neg j, m) = 0.001 \times \mathbf{0.95}$$

$$P(b, a, \neg j, m) = \mathbf{0.00095}$$

$\boldsymbol{P(b)}$	$\boldsymbol{P(\neg b)}$
0.001	0.999



$\boldsymbol{B}$	$\boldsymbol{P(a\|B)}$	$\boldsymbol{P(\neg a\|B)}$
T	0.95	0.05
F	0.29	0.71

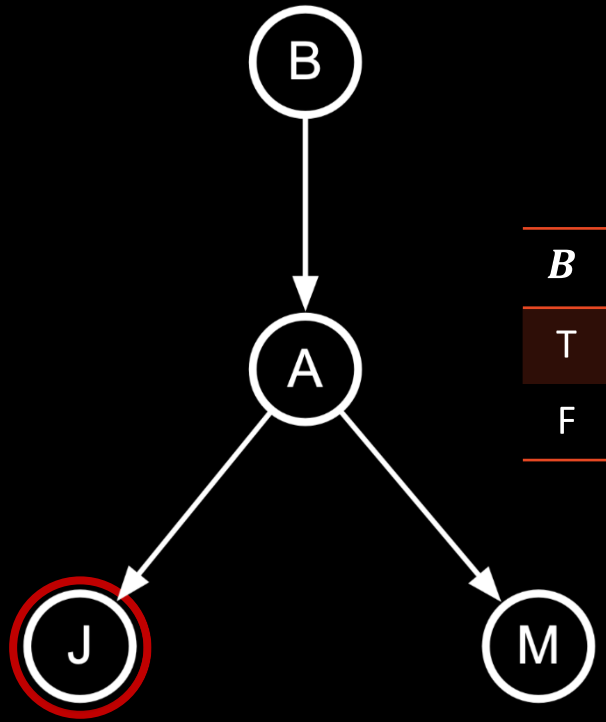
J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

$\boldsymbol{A}$	$\boldsymbol{P(j\|A)}$	$\boldsymbol{P(\neg j\|A)}$
T	0.90	0.10
F	0.05	0.95

$\boldsymbol{A}$	$\boldsymbol{P(m\|A)}$	$\boldsymbol{P(\neg m\|A)}$
T	0.70	0.30
F	0.01	0.99

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a \vert B)$	$P(\neg a \vert B)$
T	0.95	0.05
F	0.29	0.71

$$P(B, A, J, M) = P(B)\, P(A|B)\, P(J|A)\, P(M|A)$$

$$P(b, a, \neg j, m) = P(b) \times P(a|b) \times P(\neg j|a) \times P(m|a)$$
$$P(b, a, \neg j, m) = 0.001 \times 0.95 \times \mathbf{0.10}$$

$$P(b, a, \neg j, m) = \mathbf{0.000095}$$

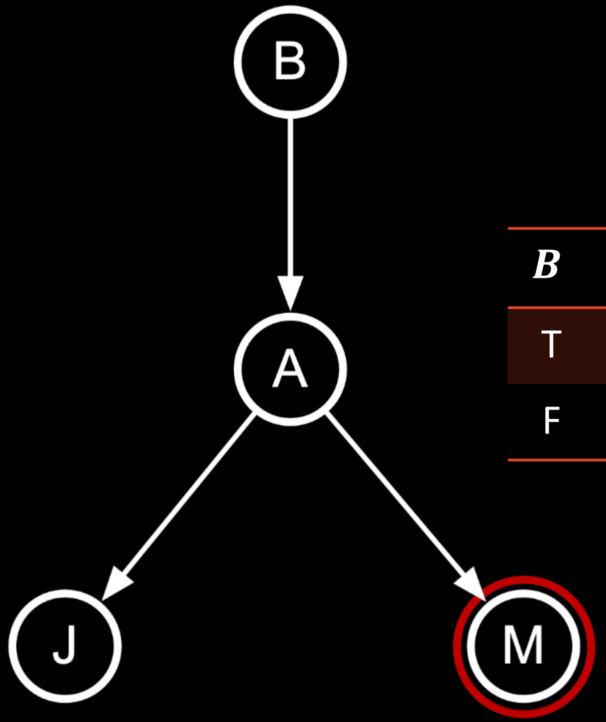
J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

A	$P(j \vert A)$	$P(\neg j \vert A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m \vert A)$	$P(\neg m \vert A)$
T	0.70	0.30
F	0.01	0.99

You are at work. You receive a phone call from your neighbors Mary and John who say that they think they hears your alarm going off. Is it possible that you are being burgled?

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a \mid B)$	$P(\neg a \mid B)$
T	0.95	0.05
F	0.29	0.71

$$P(B, A, J, M) = P(B)\,P(A|B)\,P(J|A)\,P(M|A)$$

$$P(b, a, \neg j, m) = P(b) \times P(a|b) \times P(\neg j|a) \times P(m|a)$$
$$P(b, a, \neg j, m) = 0.001 \times 0.95 \times 0.10 \times \mathbf{0.70}$$

$$P(b, a, \neg j, m) = \mathbf{0.0000665}$$

J	John calls	{T, F}
M	Mary calls	{T, F}
A	Alarm	{T, F}
B	Burglary	{T, F}

A	$P(j \mid A)$	$P(\neg j \mid A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m \mid A)$	$P(\neg m \mid A)$
T	0.70	0.30
F	0.01	0.99

Given a **joint distribution query**: $P(q_1, \dots, q_k)$

- **Evidence variables**: $None$
- **Query variable(s)**: $Q_1, \dots, Q_k$
- **Hidden variables**: $H_1, \dots, H_r$

All the variables of the model $X_1, \dots, X_n$

- **Step 1**: Calculate the joint distribution from the network using the Bayes Network rule:

$$P(h_1, \dots, h_r, q_1, \dots, q_k) = \prod_i P(x_i | parents(X_i))$$

- **Step 2**: Sum out to get the joint probability of query and evidence:

$$P(q_1, \dots, q_k) = \sum_{h_1, \dots, h_r} P(h_1, \dots, h_r, e_1, \dots, e_k)$$

$P(\neg j) = ?$

$$P(\neg j) = \sum_{B,A,M} P(B, A, \neg j, M)$$

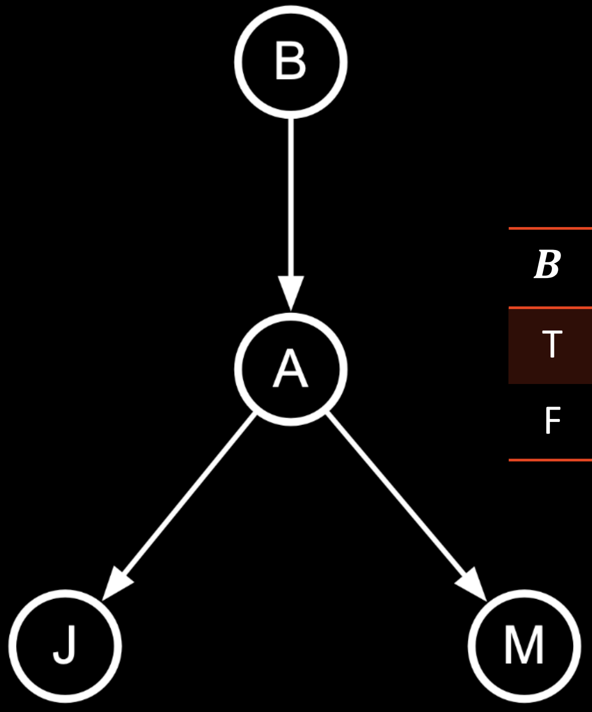
$$P(B, A, J, M) = P(B) \times P(A|B) \times P(J|A) \times P(M|A)$$

$$P(\neg j) = \sum_{B,A,M} P(B)\ P(A|B)\ P(\neg j|A)\ P(M|A)$$

$$\begin{aligned}P(\neg j) = {} & P(b)P(a|b)P(\neg j|a)P(m|a) + \\ & P(b)P(a|b)P(\neg j|a)P(\neg m|a) + \\ & P(b)P(\neg a|b)P(\neg j|\neg a)P(m|\neg a) + \\ & P(b)P(\neg a|b)P(\neg j|\neg a)P(\neg m|\neg a) + \\ & P(\neg b)P(a|\neg b)P(\neg j|a)P(m|a) + \\ & P(\neg b)P(a|\neg b)P(\neg j|a)P(\neg m|a) + \\ & P(\neg b)P(\neg a|\neg b)P(\neg j|\neg a)P(m|\neg a) + \\ & P(\neg b)P(\neg a|\neg b)P(\neg j|\neg a)P(\neg m|\neg a)\end{aligned}$$

$$\begin{aligned}P(\neg j) = {} & 0.001 \times 0.95 \times 0.1 \times 0.7 + \\ & 0.001 \times 0.95 \times 0.1 \times 0.3 + \\ & 0.001 \times 0.05 \times 0.95 \times 0.01 + \\ & 0.001 \times 0.05 \times 0.95 \times 0.99 + \\ & 0.999 \times 0.29 \times 0.1 \times 0.7 + \\ & 0.999 \times 0.29 \times 0.1 \times 0.3 + \\ & 0.999 \times 0.71 \times 0.95 \times 0.01 + \\ & 0.999 \times 0.71 \times 0.95 \times 0.99 \qquad = \mathbf{0.9775}\end{aligned}$$

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a|B)$	$P(\neg a|B)$
T	0.95	0.05
F	0.29	0.71

A	$P(j|A)$	$P(\neg j|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m|A)$	$P(\neg m|A)$
T	0.70	0.30
F	0.01	0.99

Given the joint distribution, we have a **conditional query**: $P(Q_1, \ldots, Q_l | e_1, \ldots, e_k)$

- **Evidence variables**: $E_1 = e_1, \ldots, E_k = e_k$
- **Query variable(s)**: $Q_1, \ldots, Q_l$
- **Hidden variables**: $H_1, \ldots, H_r$

All the variables of the model $X_1, \ldots, X_n$

- **Step 1:** Using the Product rule, if we calculate the joint probability instead of the conditional:

$$P(Q_1, \ldots, Q_l | e_1, \ldots, e_k) = \frac{\mathrm{P}(\mathrm{Q}_1, \ldots, \mathrm{Q}_\mathrm{l}, e_1, \ldots, e_k)}{\mathrm{P}(e_1, \ldots, e_k)}$$

- **Step 2:** Calculate the joint distribution from the network using the Bayes Network rule:

$$P(Q_1, \ldots, Q_l, h_1, \ldots, h_r, e_1, \ldots, e_k) = \prod_i P(x_i | parents(X_i))$$

- **Step 3:** Sum out to get joint of query and evidence :

$$P(Q_1, \ldots, Q_l, e_1, \ldots, e_k) = \sum_{h_1, \ldots, h_r} P(Q_1, \ldots, Q_l, h_1, \ldots, h_r, e_1, \ldots, e_k)$$

- **Step 4:** Recursively, using the same algorithm, calculate:

$$Z = P(e_1, \ldots, e_{k_i})$$

- **Step 5:** Normalize:

$$P(Q_1, \ldots, Q_l | e_1, \ldots, e_k) = \frac{1}{\mathrm{Z}} \mathrm{P}(\mathrm{Q}_1, \ldots, \mathrm{Q}_\mathrm{l}, e_1, \ldots, e_k)$$

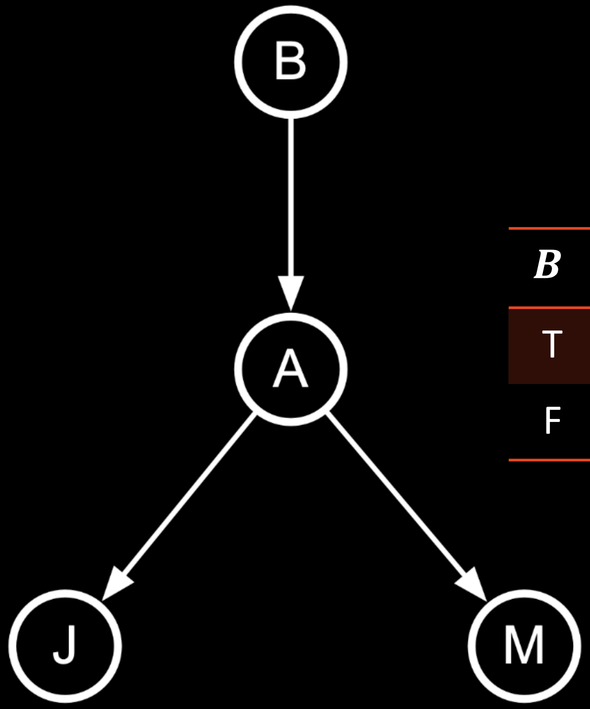
$$P(a|\neg j) = \frac{P(a, \neg j)}{P(\neg j)}$$

$$P(a, \neg j) = \sum_{B,M} P(B)\; P(a|B)\; P(\neg j|a)\; P(M|a)$$

$$\begin{aligned} P(a, \neg j) = {} & P(b)P(a|b)P(\neg j|a)P(m|a) + \\ & P(b)P(a|b)P(\neg j|a)P(\neg m|a) + \\ & P(\neg b)P(a|\neg b)P(\neg j|a)P(m|a) + \\ & P(\neg b)P(a|\neg b)P(\neg j|a)P(\neg m|a) \end{aligned}$$

$$\begin{aligned} P(a, \neg j) = {} & 0.001 \times 0.95 \times 0.1 \times 0.3 + \\ & 0.001 \times 0.95 \times 0.1 \times 0.7 + \\ & 0.999 \times 0.29 \times 0.1 \times 0.3 + \\ & 0.999 \times 0.29 \times 0.1 \times 0.7 = \quad \mathbf{0.0291} \end{aligned}$$

$P(b)$	$P(\neg b)$
0.001	0.999



B	$P(a\|B)$	$P(\neg a\|B)$
T	0.95	0.05
F	0.29	0.71

A	$P(j\|A)$	$P(\neg j\|A)$
T	0.90	0.10
F	0.05	0.95

A	$P(m\|A)$	$P(\neg m\|A)$
T	0.70	0.30
F	0.01	0.99

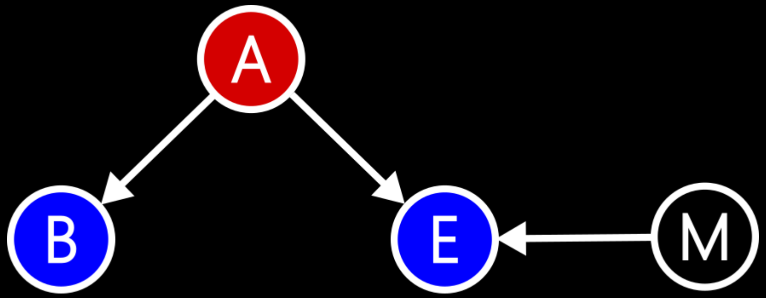
Given the joint distribution, we have a **conditional independence query**:

$$\{Q_1, \dots, Q_m\} \perp \{Q'_1, \dots, Q'_j\} \mid \{E_1, \dots, E_k\}$$

$$\left.\begin{array}{l} \bullet\ \textbf{Evidence variables}: E_1 = e_1, \dots, E_k = e_k \\ \bullet\ \textbf{Query variable(s)}: Q_1, \dots, Q_l \\ \bullet\ \textbf{Hidden variables}: H_1, \dots, H_r \end{array}\right\} \text{All the variables of the model } X_1, \dots, X_n$$

$$P(B,E|A) = \frac{1}{P(A)} \sum_M P(B,E,M,A)$$
$$= \frac{1}{P(A)} \sum_M P(A)\, P(M)\, P(E|M,A)\, P(B|A)$$
$$= P(B|A) \sum_M P(E|M,A)\, P(M)$$

$$P(E\,|A) = \frac{1}{P(A)} \sum_{B,A,M} P(B,E,M,A)$$
$$= \frac{1}{P(A)} \sum_{B,A,M} P(A)\, P(M)\, P(E|M,A)\, P(B|A)$$
$$= \sum_B P(B|A)\, \sum_M P(E|M,A)\, P(M)$$
$$= \sum_M P(E|M,A)\, P(M)$$



$$P(B,E|A) = \mathrm{P}(B|A)\, \mathrm{P}(E|A)$$

So, **yes**, B and E are conditionally independent given A.

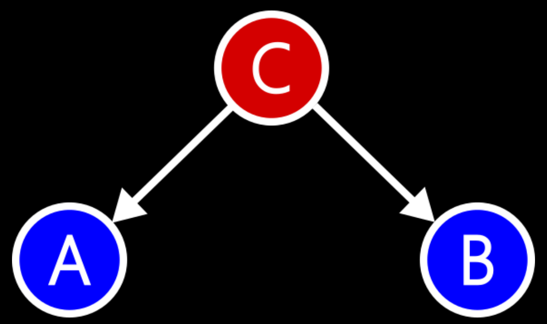
The **D-separation algorithm**, proposed by Pearl in the 1980s, can automatically discover, with some limitations, if variables are conditionally independent.

Tail-Tail rule (or Common cause)

It's possible to demonstrate that in this simple case: $A \perp B \mid C$:

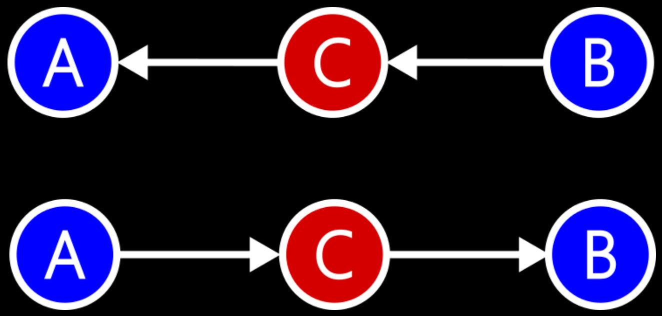
$$P(A,B,C) = P(C)\, P(A|C)\, P(B|C)$$

$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = \boldsymbol{P(A|C)\, P(B|C)}$$



Head-Tail rule (or Chain)

It's possible to demonstrate that in this simple case: $A \perp B \mid C$:



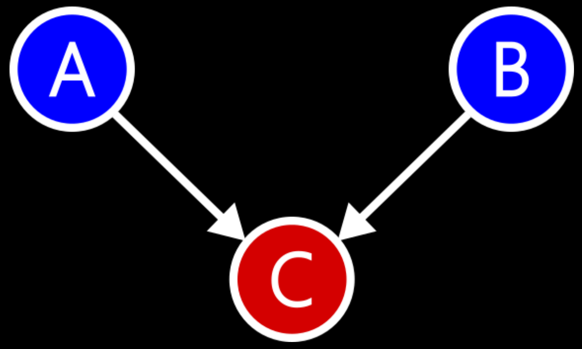
$$P(A,B,C) = P(A)\,P(C|A)P\,(B|C)$$
$$= P(A,C)\,P(B|C)$$
$$= P(A|C)\,P(C)\,P(B|C)$$

$$P(A,B|C) = \frac{P(A,B,C)}{P(C)} = \boldsymbol{P(A|C)\,P(B|C)}$$

Head-Head rule (or Collider)

It's possible to demonstrate that in this simple case: $A \perp B \mid C$:

$$P(A,B,C) = P(A)\,P(B)\,P(C|A,B)$$



$$P(A,B|C) = \frac{P(A,B,C)}{P(C)}$$

We don't know

It's not always the case: it depends how C is behaves. For example, knowing A and C also gives you information about B:

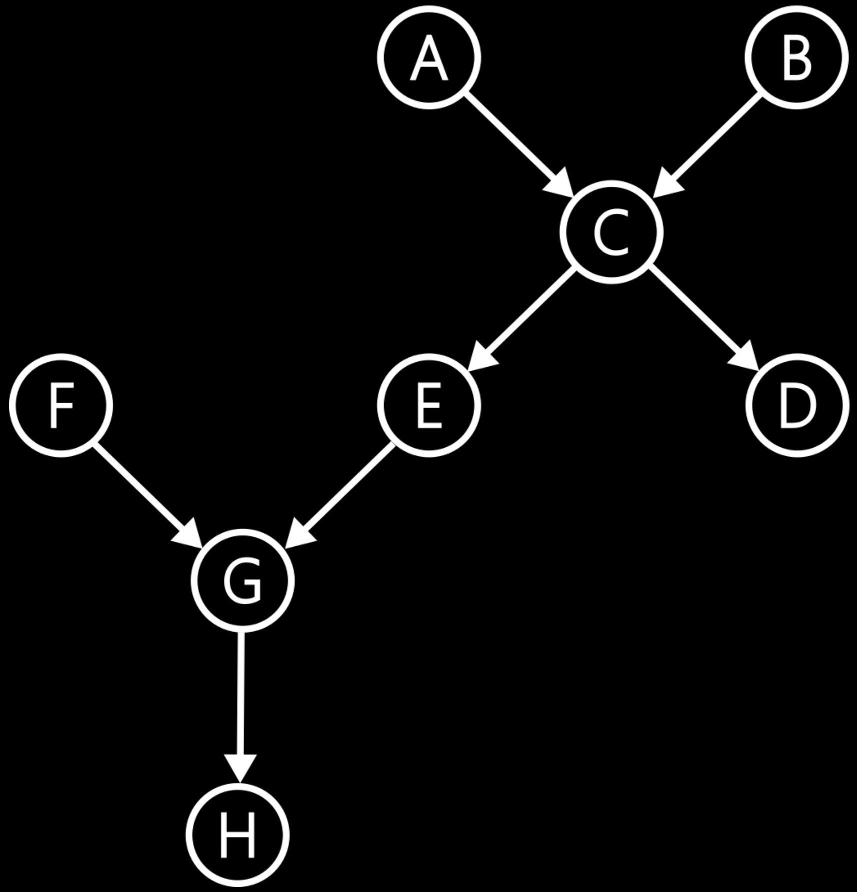
$$C = \begin{cases} 1 & if\ A = B \\ 0 & otherwise \end{cases}$$

- Given the query $\{A_1, \ldots, A_m\} \perp \{B_1, \ldots, B_l\} \mid \{C_1, \ldots, C_k\}$

- We define a path between vertices of A and B as **blocked** if it passes through a vertex c is one of these two conditions happen:

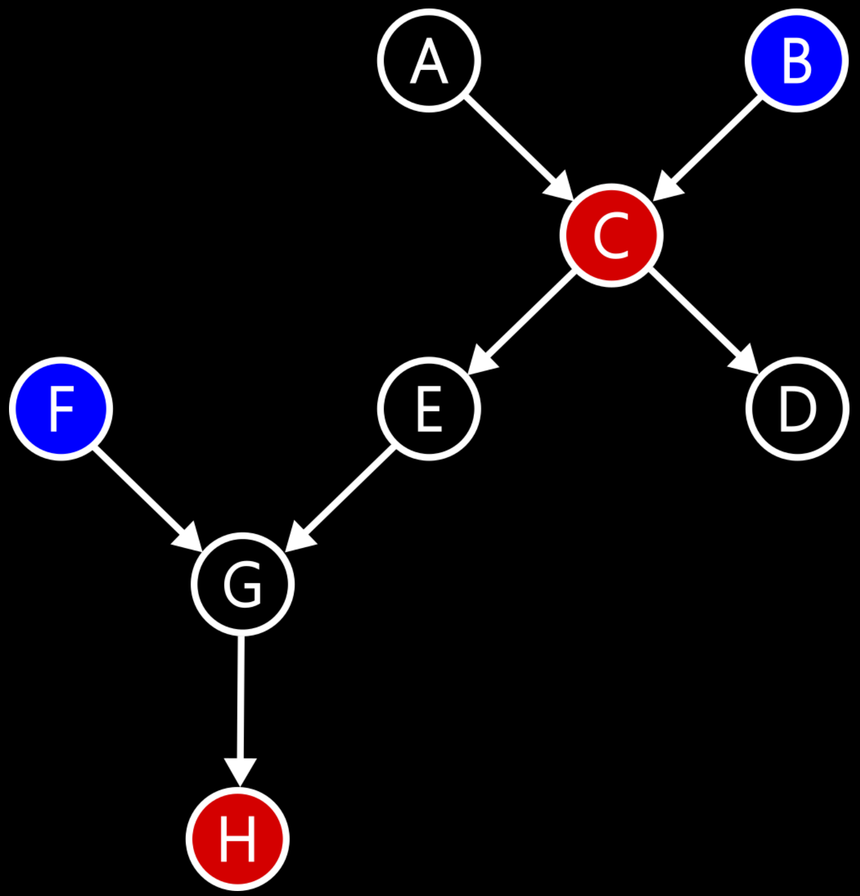
  - the edges are **head-tail** or **tail-tail** and $c \in C$
  - the edges are **head-head** and $c \notin C$ and none of the descendants belong to C

- If all such paths are blocked, then A and B are **D-separated** by C and therefore **conditionally independent** with respect to C



Are the random variables D-separable and therefore conditionally independent?

1. $B \perp F \mid C, H$

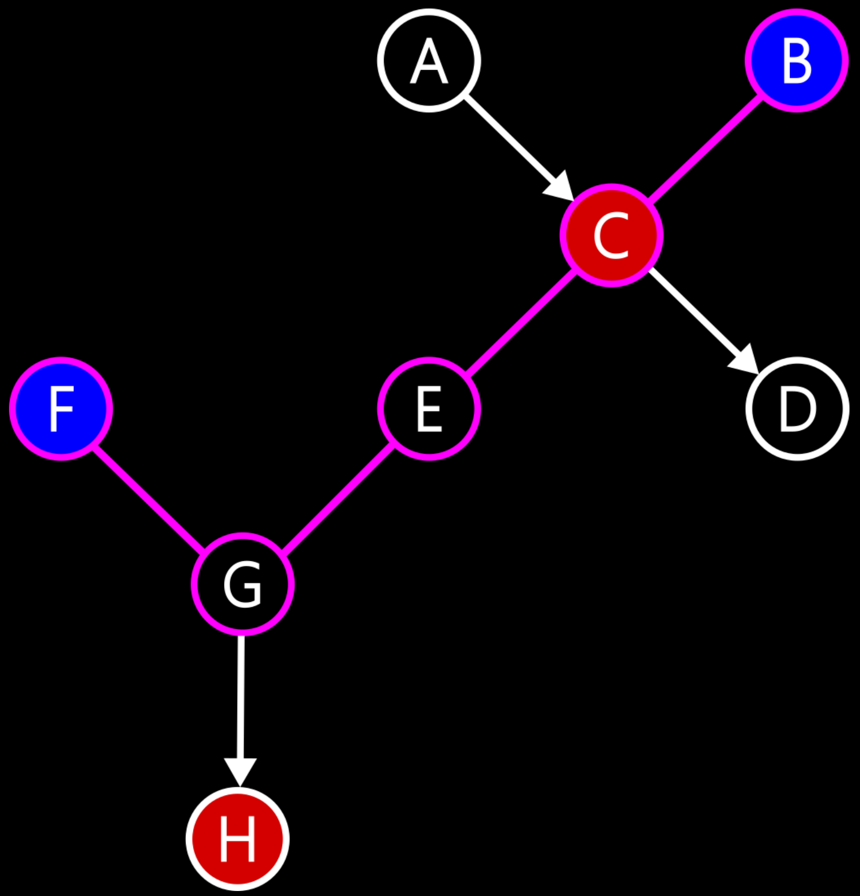


Are the random variables D-separable and therefore conditionally independent?

1. $B \perp F \mid C, H$

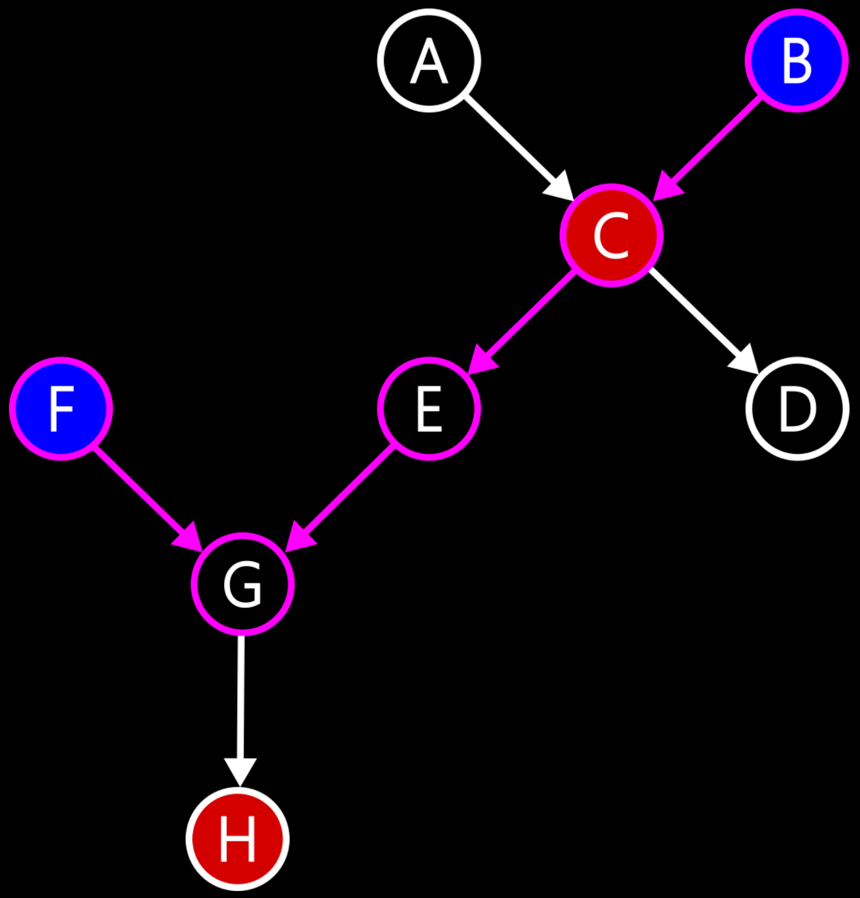
D-separation

EXAMPLE



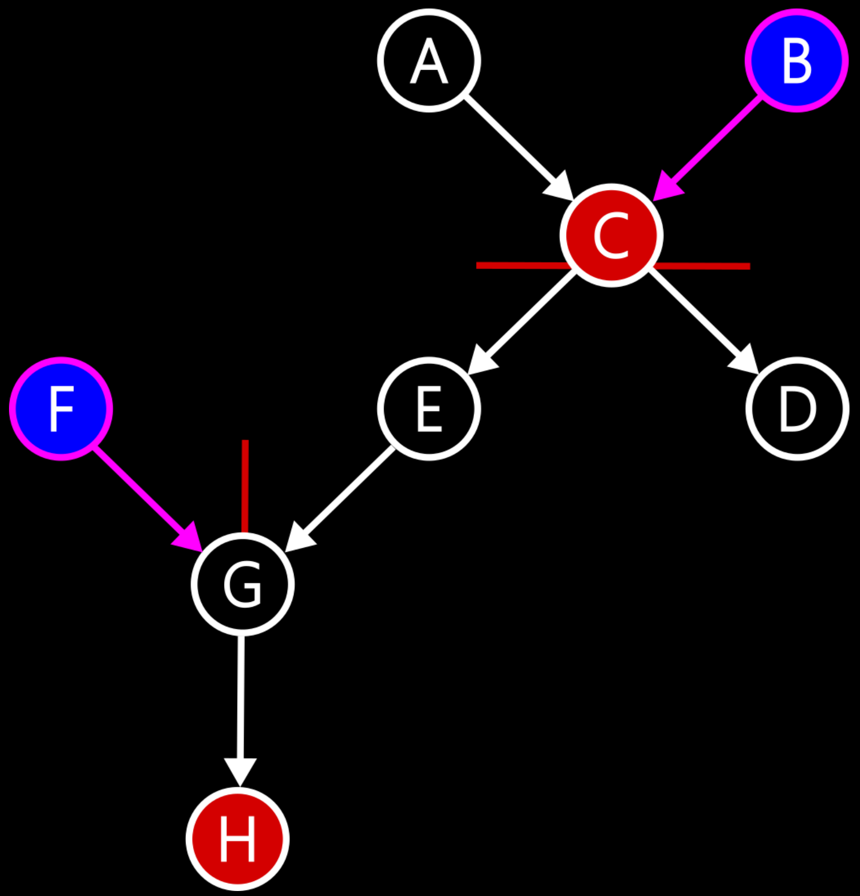
Are the random variables D-separable and therefore conditionally independent?

1. $B \perp F \mid C, H$



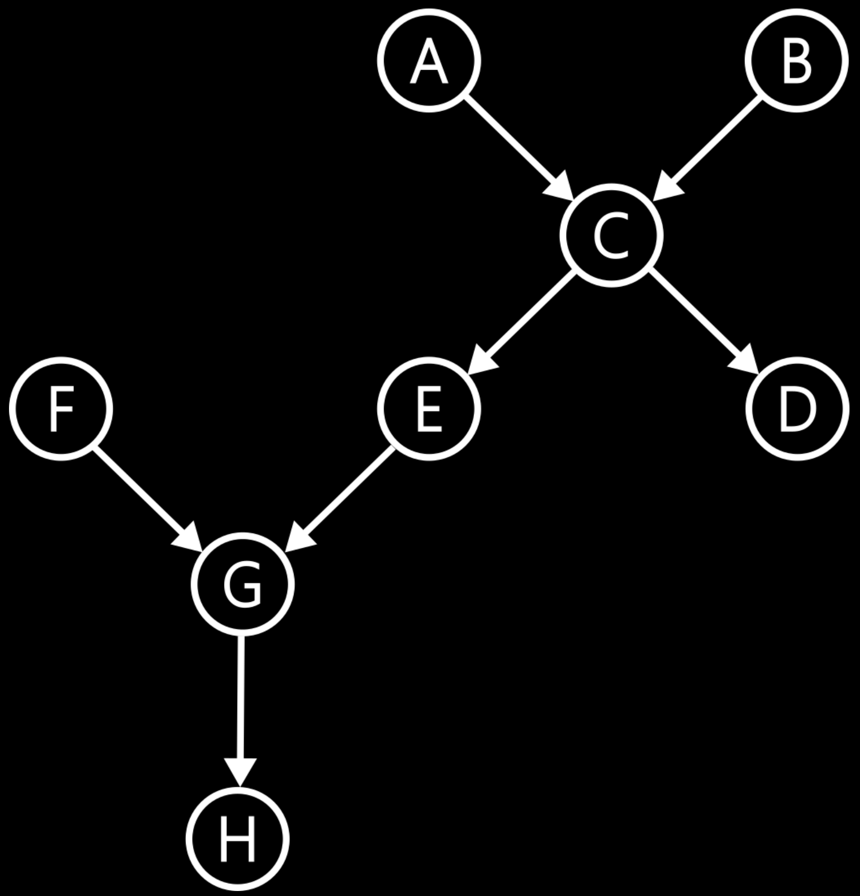
Are the random variables D-separable and therefore conditionally independent?

1. $B \perp F \mid C, H$



Are the random variables D-separable and therefore conditionally independent?

$1.\ B \perp F \mid C, H$ YES C. I.



Are the random variables D-separable and therefore conditionally independent?

1. $B \perp F \mid C, H$	YES	C. I.
2. $A \perp E \mid C$	YES	C. I.
3. $E \perp G \mid C$	NO	Unknown
4. $E \perp D \mid H$	NO	Unknown
5. $F \perp E \mid H$	NO	Unknown
6. $A \perp B \mid H$	NO	Unknown
7. $A \perp B \mid F$	YES	C. I.

READINGS

Chapter 14

QUESTIONS ?

ARTIFICIAL INTELLIGENCE

COMP 131

FABRIZIO SANTINI

VERSION 4.2