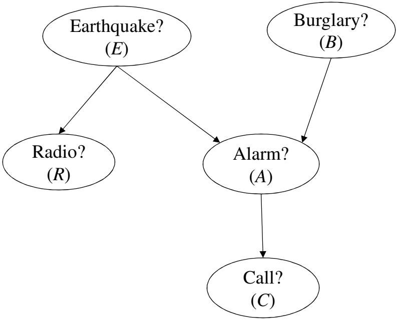
Chapter 4: Bayesian Networks

Adnan Darwiche[1]

October 6, 2009

[1]Lecture slides for *Modeling and Reasoning with Bayesian Networks*, Adnan Darwiche, Cambridge University Press, 2009.

Capturing Independence Graphically

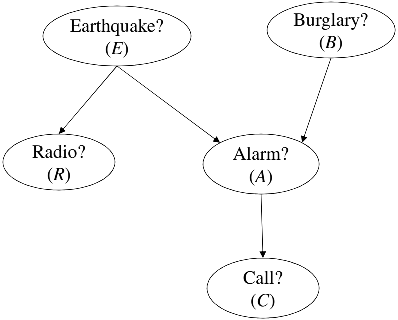


Assume that edges in this graph represent direct causal influences among these variables.

Example

The alarm triggering (A) is a direct cause of receiving a call from a neighbor (C).

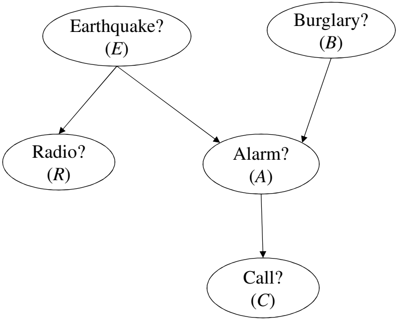
Capturing Independence Graphically



We expect our belief in C to be influenced by evidence on R

If we get a radio report that an earthquake took place in our neighborhood, our belief in the alarm triggering would probably increase, which would also increase our belief in receiving a call from our neighbor.

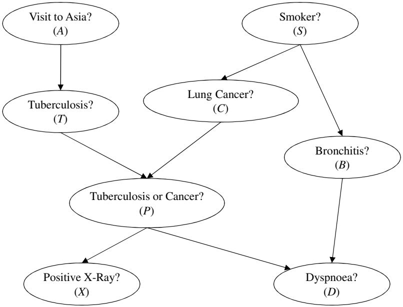
Capturing Independence Graphically



We would not change this belief, however, if we knew for sure that the alarm did not trigger.

C independent of R given $\neg A$

Capturing Independence Graphically



We would clearly find a visit to Asia relevant to our belief in the X-Ray test coming out positive, but we would find the visit irrelevant if we know for sure that the patient does not have Tuberculosis.

X is dependent on A, but is independent of A given $\neg T$

Capturing Independence Graphically

These examples of independence are all implied by a formal interpretation of each DAG as a set of conditional independence statements.

Capturing Independence Graphically

These examples of independence are all implied by a formal interpretation of each DAG as a set of conditional independence statements.

$\text{Parents}(V)$

variables N with an edge from N to V

Capturing Independence Graphically

These examples of independence are all implied by a formal interpretation of each DAG as a set of conditional independence statements.

Parents(V)

variables N with an edge from N to V

Descendants(V)

variables N with a directed path from V to N.
V is said to be an ancestor of N

Capturing Independence Graphically

These examples of independence are all implied by a formal interpretation of each DAG as a set of conditional independence statements.

Parents(V)

variables N with an edge from N to V

Descendants(V)

variables N with a directed path from V to N.
V is said to be an ancestor of N

Non_Descendants(V)

variables other than V, Parents(V) and Descendants(V)

Capturing Independence Graphically

Markovian assumptions of a DAG

We will formally interpret each DAG G as a compact representation of the following independence statements, denoted $\mathrm{Markov}(G)$:

$$I(V, \mathrm{Parents}(V), \mathrm{Non\_Descendants}(V)),$$

for all variables V in DAG G

Capturing Independence Graphically

Markovian assumptions of a DAG

We will formally interpret each DAG G as a compact representation of the following independence statements, denoted $\mathrm{Markov}(G)$:

$$I(V, \mathrm{Parents}(V), \mathrm{Non\_Descendants}(V)),$$

for all variables V in DAG G

DAG as a causal structure

$\mathrm{Parents}(V)$ denote the direct causes of V and $\mathrm{Descendants}(V)$ denote the effects of V

Capturing Independence Graphically

Markovian assumptions of a DAG

We will formally interpret each DAG G as a compact representation of the following independence statements, denoted $\mathrm{Markov}(G)$:

$$I(V, \mathrm{Parents}(V), \mathrm{Non\_Descendants}(V)),$$

for all variables V in DAG G

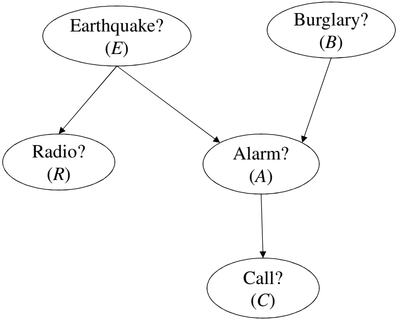
DAG as a causal structure

$\mathrm{Parents}(V)$ denote the direct causes of V and $\mathrm{Descendants}(V)$ denote the effects of V

Markovian assumptions restated

Given the direct causes of a variable, our beliefs in that variable become independent of its non–effects.

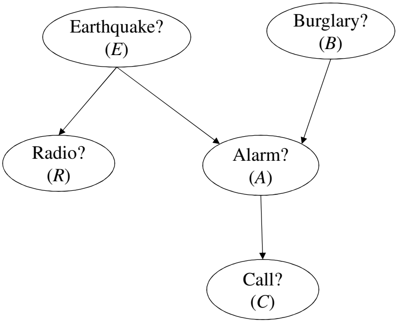
Capturing Independence Graphically



Markovian assumptions, Markov(G):

$I(C,$

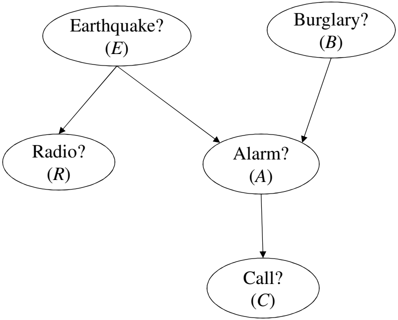
Capturing Independence Graphically



Markovian assumptions, $\text{Markov}(G)$:

$$I(C, A, \{B, E, R\})$$

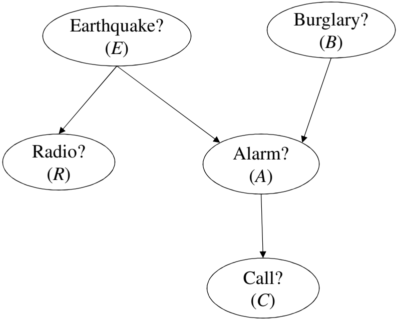
Capturing Independence Graphically



Markovian assumptions, Markov(G):

$I(C, A, \{B, E, R\})$
$I(R,$

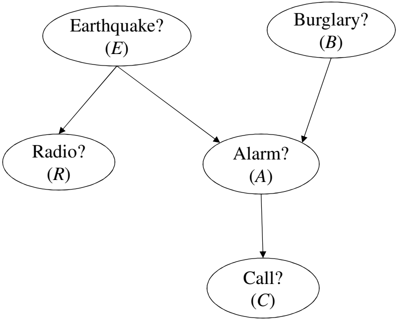
Capturing Independence Graphically



Markovian assumptions, $\text{Markov}(G)$:

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$

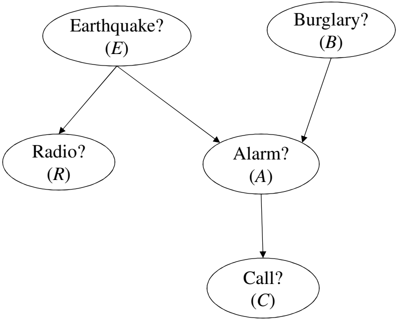
Capturing Independence Graphically



Markovian assumptions, $\mathrm{Markov}(G)$:

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A,$$

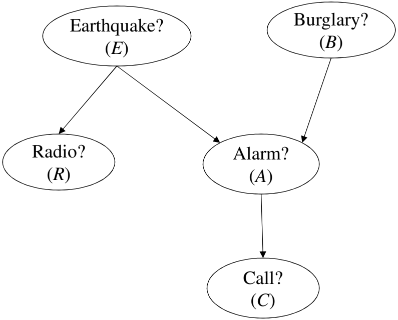
Capturing Independence Graphically



Markovian assumptions, $\mathrm{Markov}(G)$:

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A, \{B, E\}, R)$$

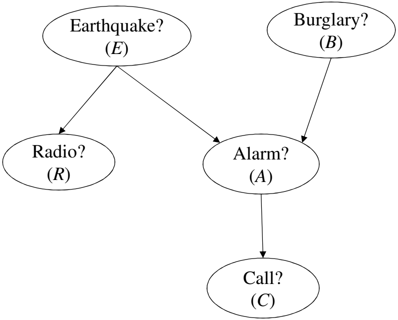
Capturing Independence Graphically



Markovian assumptions, Markov(G):

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A, \{B, E\}, R)$$
$$I(B,$$

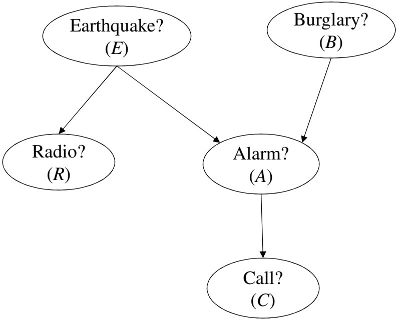
Capturing Independence Graphically



Markovian assumptions, $\mathrm{Markov}(G)$:

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A, \{B, E\}, R)$$
$$I(B, \emptyset, \{E, R\})$$

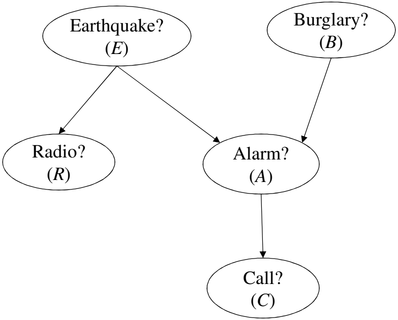
Capturing Independence Graphically



Markovian assumptions, Markov(G):

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A, \{B, E\}, R)$$
$$I(B, \emptyset, \{E, R\})$$
$$I(E,$$

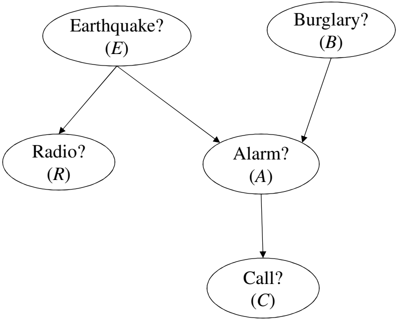
Capturing Independence Graphically



Markovian assumptions, $\text{Markov}(G)$:

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A, \{B, E\}, R)$$
$$I(B, \emptyset, \{E, R\})$$
$$I(E, \emptyset, B)$$

Capturing Independence Graphically



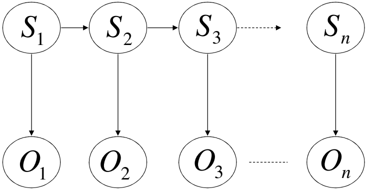
Markovian assumptions, Markov(G):

$$I(C, A, \{B, E, R\})$$
$$I(R, E, \{A, B, C\})$$
$$I(A, \{B, E\}, R)$$
$$I(B, \emptyset, \{E, R\})$$
$$I(E, \emptyset, B)$$

Variables B and E have no parents, hence, they are marginally independent of their non-descendants.

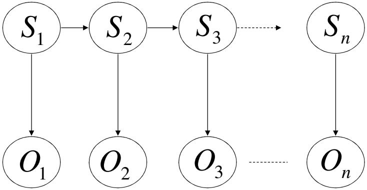
Capturing Independence Graphically

Hidden Markov Model



Capturing Independence Graphically

Hidden Markov Model

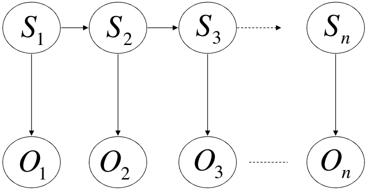


$S_1, S_2, \ldots, S_n$

The state of a dynamic system at time points $1, 2, \ldots, n$

Capturing Independence Graphically

Hidden Markov Model



$S_1, S_2, \ldots, S_n$

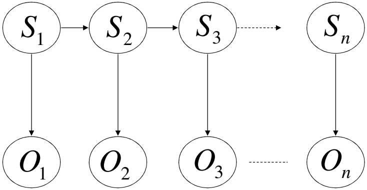
The state of a dynamic system at time points $1, 2, \ldots, n$

$O_1, O_2, \ldots, O_n$

Sensors that measure the system state at the corresponding time points.

Capturing Independence Graphically

Hidden Markov Model



$S_1, S_2, \ldots, S_n$

The state of a dynamic system at time points $1, 2, \ldots, n$

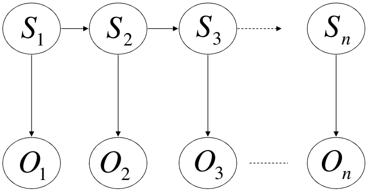
$O_1, O_2, \ldots, O_n$

Sensors that measure the system state at the corresponding time points.

Usually, one has some information about the sensor readings and is interested in computing beliefs in the system states.

Capturing Independence Graphically

Hidden Markov Model

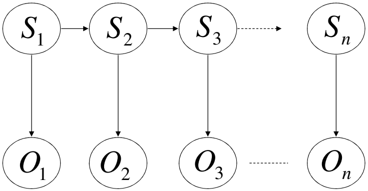


The Markovian assumptions imply that

once we know the state of the system at the previous time point, $t-1$, our belief in the present system state, at t, is no longer influenced by any other information about the past.

Capturing Independence Graphically

Hidden Markov Model



The Markovian assumptions imply that

once we know the state of the system at the previous time point, $t-1$, our belief in the present system state, at t, is no longer influenced by any other information about the past.

Characteristic property of HMMs

$$I(S_t, \{S_{t-1}\}, \{S_1, \ldots, S_{t-2}, O_1, \ldots, O_{t-1}\})$$

Capturing Independence Graphically

Interpretation of DAGs in terms of conditional independence makes no reference to causality

even though we have used causality to motivate this interpretation.

Capturing Independence Graphically

Interpretation of DAGs in terms of conditional independence makes no reference to causality

even though we have used causality to motivate this interpretation.

If one constructs the DAG based on causal perceptions

one tends to agree with the independencies declared by the DAG.

Capturing Independence Graphically

Interpretation of DAGs in terms of conditional independence makes no reference to causality

even though we have used causality to motivate this interpretation.

If one constructs the DAG based on causal perceptions

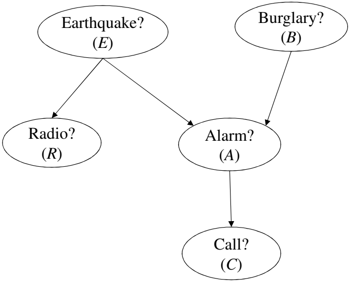
one tends to agree with the independencies declared by the DAG.

Possible to have a DAG that does not match our causal perceptions

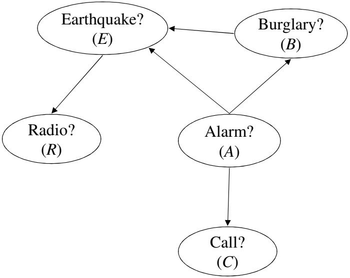
yet we agree with the independencies declared by the DAG.

Capturing Independence Graphically

DAG is causal

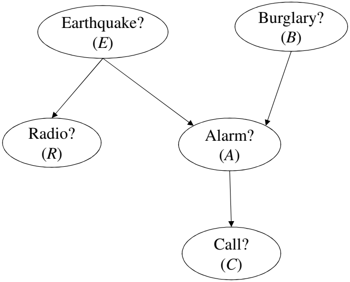


DAG is not causal

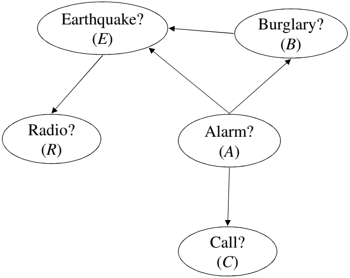


Capturing Independence Graphically

DAG is causal

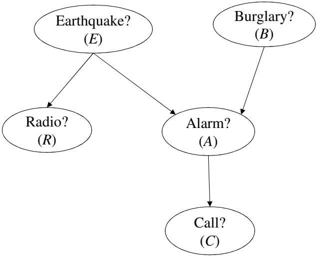


DAG is not causal



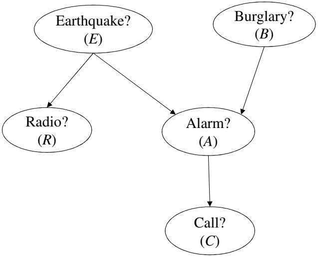
Every independence which is declared (or implied) by the second DAG is also declared (or implied) by the first one. Hence, if we accept the first DAG, then we must also accept the second.

Parameterizing the Independence Structure



DAG G is a partial specification of our state of belief Pr

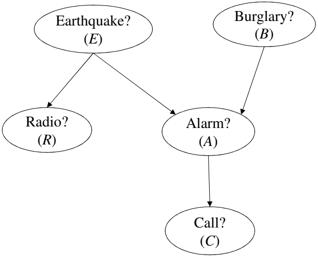
Parameterizing the Independence Structure



DAG G is a partial specification of our state of belief Pr

By constructing G, we are saying that Pr must satisfy the independence assumptions in $\mathrm{Markov}(G)$

Parameterizing the Independence Structure

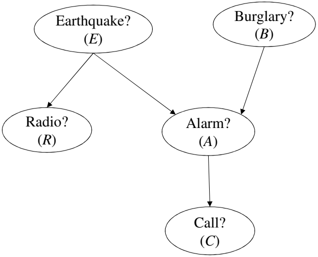


DAG G is a partial specification of our state of belief Pr

By constructing G, we are saying that Pr must satisfy the independence assumptions in $\mathrm{Markov}(G)$

This constrains Pr but does not uniquely define it.

Parameterizing the Independence Structure



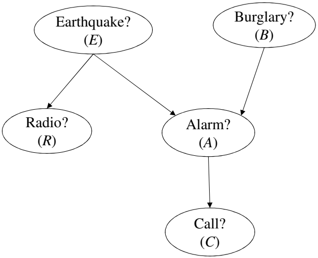
DAG G is a partial specification of our state of belief Pr

By constructing G, we are saying that Pr must satisfy the independence assumptions in $\mathrm{Markov}(G)$

This constrains Pr but does not uniquely define it.

We can augment the DAG G by a set of conditional probabilities that together with $\mathrm{Markov}(G)$ define the distribution Pr uniquely.

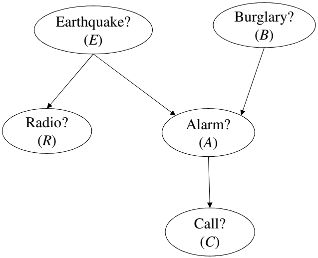
Parameterizing the Independence Structure



For every variable X and its parents $\mathbf{U}$

Need probability $\Pr(x|\mathbf{u})$ for every value x and every instantiation $\mathbf{u}$

Parameterizing the Independence Structure



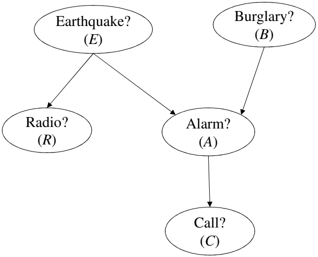
For every variable X and its parents $\mathbf{U}$

Need probability $\Pr(x|\mathbf{u})$ for every value x and every instantiation $\mathbf{u}$

We need to provide the following conditional probabilities

$\Pr(c|a)$, $\Pr(r|e)$, $\Pr(a|b,e)$, $\Pr(e)$, $\Pr(b)$,
where a, b, c, e and r are values of variables A, B, C, E and R

Parameterizing the Independence Structure

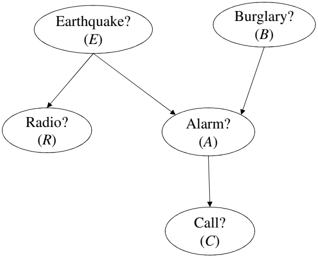


Conditional probabilities for variable C

A	C	$\Pr(c\|a)$
true	true	.80
true	false	.20
false	true	.001
false	false	.999

Conditional Probability Table (CPT)

Parameterizing the Independence Structure



Conditional probabilities for variable C

A	C	$\Pr(c\|a)$
true	true	.80
true	false	.20
false	true	.001
false	false	.999

Conditional Probability Table (CPT)

$$\Pr(c|a) + \Pr(\bar{c}|a) = 1 \text{ and } \Pr(c|\bar{a}) + \Pr(\bar{c}|\bar{a}) = 1$$

Two of the probabilities in the above CPT are redundant and can be inferred from the other two. We only need 10 independent probabilities to completely specify the CPTs for this DAG.

Definition

A Bayesian network for variables $\mathbf{Z}$ is a pair (G, Θ), where

- G is a directed acyclic graph over variables $\mathbf{Z}$, called the network structure.
- Θ is a set of conditional probability tables (CPTs), one for each variable in $\mathbf{Z}$, called the network parametrization.

Definition

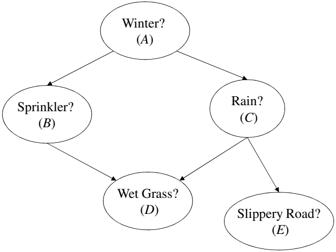
A Bayesian network for variables $\mathbf{Z}$ is a pair (G, Θ), where

- G is a directed acyclic graph over variables $\mathbf{Z}$, called the network structure.
- Θ is a set of conditional probability tables (CPTs), one for each variable in $\mathbf{Z}$, called the network parametrization.

- $\Theta_{X|\mathbf{U}}$: CPT for variable X and its parents $\mathbf{U}$
- $X\mathbf{U}$: called a network family
- $\theta_{x|\mathbf{u}} = \Pr(x|\mathbf{u})$: called a network parameter

We must have $\sum_x \theta_{x|\mathbf{u}} = 1$ for every parent instantiation $\mathbf{u}$

An Example Bayesian Network



A	B	$\Theta_{B\|A}$
true	true	.2
true	false	.8
false	true	.75
false	false	.25

A	C	$\Theta_{C\|A}$
true	true	.8
true	false	.2
false	true	.1
false	false	.9

A	Θ_A
true	.6
false	.4

B	C	D	$\Theta_{D\|B,C}$
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

C	E	$\Theta_{E\|C}$
true	true	.7
true	false	.3
false	true	0
false	false	1

Notation

A network instantiation

is an instantiation of all network variables.

Notation

A network instantiation

is an instantiation of all network variables.

$\theta_{x|\mathbf{u}} \sim \mathbf{z}$

means that instantiations $x\mathbf{u}$ and $\mathbf{z}$ are compatible (i.e., agree on the values they assign to their common variables).

Notation

A network instantiation

is an instantiation of all network variables.

$\theta_{x|\mathbf{u}} \sim \mathbf{z}$

means that instantiations $x\mathbf{u}$ and $\mathbf{z}$ are compatible (i.e., agree on the values they assign to their common variables).

Example

θ_a, $\theta_{b|a}$, $\theta_{\bar{c}|a}$, $\theta_{d|b,\bar{c}}$, and $\theta_{\bar{e}|\bar{c}}$ are all the network parameters compatible with network instantiation $a, b, \bar{c}, d, \bar{e}$

The Distribution of a Bayesian Network

A Bayesian network induces distribution

$$\Pr(\mathbf{z}) \stackrel{\mathit{def}}{=} \prod_{\theta_{x|\mathbf{u}} \sim \mathbf{z}} \theta_{x|\mathbf{u}}$$

The Distribution of a Bayesian Network

A Bayesian network induces distribution

$$\Pr(\mathbf{z}) \stackrel{def}{=} \prod_{\theta_{x|\mathbf{u}} \sim \mathbf{z}} \theta_{x|\mathbf{u}}$$

The probability assigned to a network instantiation $\mathbf{z}$

is the product of all network parameters that are compatible with $\mathbf{z}$

The Distribution of a Bayesian Network

A Bayesian network induces distribution

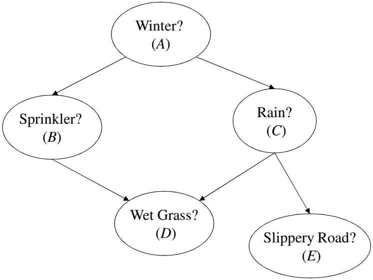
$$\Pr(\mathbf{z}) \stackrel{def}{=} \prod_{\theta_{x|\mathbf{u}} \sim \mathbf{z}} \theta_{x|\mathbf{u}}$$

The probability assigned to a network instantiation $\mathbf{z}$

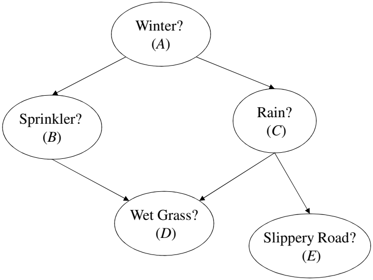
is the product of all network parameters that are compatible with $\mathbf{z}$

This is called the chain rule of Bayesian networks.

The Distribution of a Bayesian Network



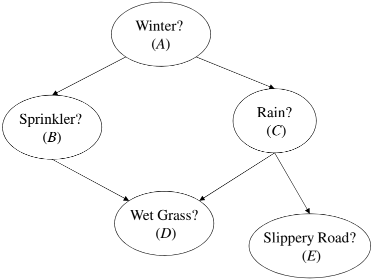
The Distribution of a Bayesian Network



$\Pr(a, b, \bar{c}, d, \bar{e})$

$$
\begin{aligned}
&= \theta_a \; \theta_{b|a} \; \theta_{\bar{c}|a} \; \theta_{d|b,\bar{c}} \; \theta_{\bar{e}|\bar{c}} \\
&= (.6)(.2)(.2)(.9)(1) \\
&= .0216
\end{aligned}
$$

The Distribution of a Bayesian Network



$\Pr(a, b, \bar{c}, d, \bar{e})$

$$
\begin{aligned}
&= \theta_a \; \theta_{b|a} \; \theta_{\bar{c}|a} \; \theta_{d|b,\bar{c}} \; \theta_{\bar{e}|\bar{c}} \\
&= (.6)(.2)(.2)(.9)(1) \\
&= .0216
\end{aligned}
$$

$\Pr(\bar{a}, \bar{b}, \bar{c}, \bar{d}, \bar{e})$

$$
\begin{aligned}
&= \theta_{\bar{a}} \; \theta_{\bar{b}|\bar{a}} \; \theta_{\bar{c}|\bar{a}} \; \theta_{\bar{d}|\bar{b},\bar{c}} \; \theta_{\bar{e}|\bar{c}} \\
&= (.4)(.25)(.9)(1)(1) \\
&= .09
\end{aligned}
$$

The Size of a Bayesian Network

The CPT $\Theta_{X|\mathbf{U}}$ is exponential in the number of parents $\mathbf{U}$

The Size of a Bayesian Network

The CPT $\Theta_{X|\mathbf{U}}$ is exponential in the number of parents $\mathbf{U}$

If every variable has d values and at most k parents

the size of any CPT is bounded by $O(d^{k+1})$

The Size of a Bayesian Network

The CPT $\Theta_{X|\mathbf{U}}$ is exponential in the number of parents $\mathbf{U}$

If every variable has d values and at most k parents

the size of any CPT is bounded by $O(d^{k+1})$

If we have n network variables

total number of network parameters is bounded by $O(n \cdot d^{k+1})$

The Size of a Bayesian Network

The CPT $\Theta_{X|\mathbf{U}}$ is exponential in the number of parents $\mathbf{U}$

If every variable has d values and at most k parents

the size of any CPT is bounded by $O(d^{k+1})$

If we have n network variables

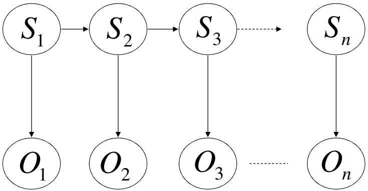
total number of network parameters is bounded by $O(n \cdot d^{k+1})$

This number is quite reasonable

as long as the number of parents per variable is relatively small.

Variable S_i has m values and similarly for variables O_i

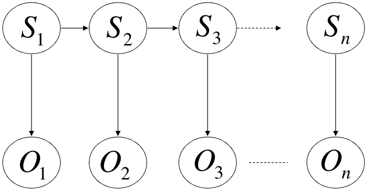
Hidden Markov Model



The Size of a Bayesian Network

Variable S_i has m values and similarly for variables O_i

Hidden Markov Model

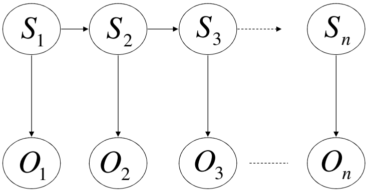


The CPT for any state variable S_i, $i > 1$, has m^2 parameters, known as transition probabilities.

The Size of a Bayesian Network

Variable S_i has m values and similarly for variables O_i

Hidden Markov Model



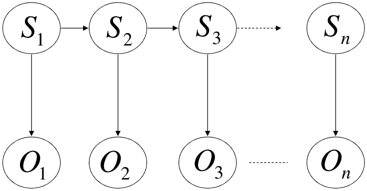
The CPT for any state variable S_i, $i > 1$, has m^2 parameters, known as transition probabilities.

The CPT for any sensor variable O_i has m^2 parameters, known as emission or sensor probabilities.

The Size of a Bayesian Network

Variable S_i has m values and similarly for variables O_i

Hidden Markov Model



The CPT for any state variable S_i, $i > 1$, has m^2 parameters, known as transition probabilities.

The CPT for any sensor variable O_i has m^2 parameters, known as emission or sensor probabilities.

The CPT for S_1 has m parameters.

A Graphical Test of Independence

The inferential power of the graphoid axioms can be captured using a graphical test, known as d-separation, which allows one to mechanically derive the independencies implied by these axioms.

A Graphical Test of Independence

The inferential power of the graphoid axioms can be captured using a graphical test, known as d-separation, which allows one to mechanically derive the independencies implied by these axioms.

$\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

iff every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

A Graphical Test of Independence

The inferential power of the graphoid axioms can be captured using a graphical test, known as d-separation, which allows one to mechanically derive the independencies implied by these axioms.

$\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

iff every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

The definition of d-separation relies on

the notion of blocking a path by a set of variables $\mathbf{Z}$

A Graphical Test of Independence

The inferential power of the graphoid axioms can be captured using a graphical test, known as d-separation, which allows one to mechanically derive the independencies implied by these axioms.

$\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

iff every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

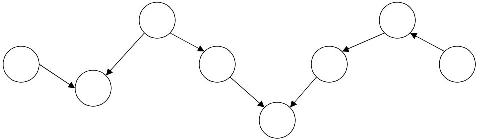
The definition of d-separation relies on

the notion of blocking a path by a set of variables $\mathbf{Z}$

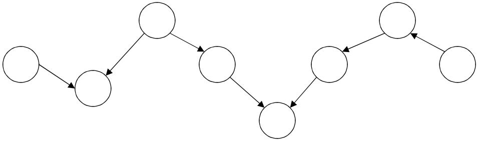
$\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ implies $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

for every probability distribution $\Pr$ induced by G

d-separation



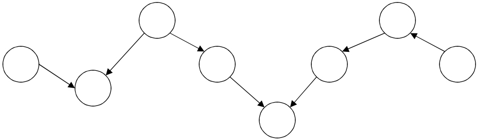
d-separation



View the path as a pipe

and view each variable W on the path as a valve.

d-separation



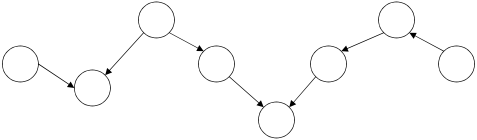
View the path as a pipe

and view each variable W on the path as a valve.

A valve W is either open or closed

depending on some conditions that we state later.

d-separation



View the path as a pipe

and view each variable W on the path as a valve.

A valve W is either open or closed

depending on some conditions that we state later.

If at least one of the valves on the path is closed

the whole path is blocked. Otherwise, the path is not blocked.

d-separation

The type of a valve

is determined by its relationship to its neighbors on the path.

d-separation

The type of a valve

is determined by its relationship to its neighbors on the path.

sequential $\longrightarrow W \longrightarrow$

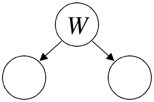
d-separation

The type of a valve

is determined by its relationship to its neighbors on the path.

sequential $\rightarrow W \rightarrow$

divergent $\leftarrow W \rightarrow$



d-separation

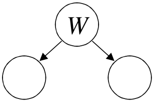
The type of a valve

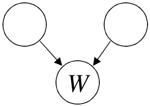
is determined by its relationship to its neighbors on the path.

sequential $\rightarrow W \rightarrow$

divergent $\leftarrow W \rightarrow$

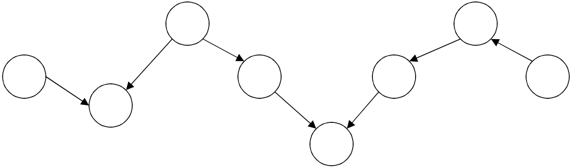
convergent $\rightarrow W \leftarrow$





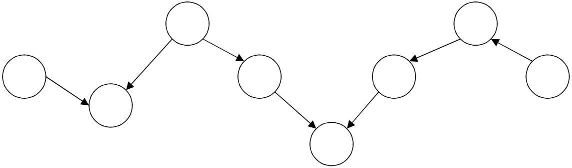
d-separation

A path with 6 valves



d-separation

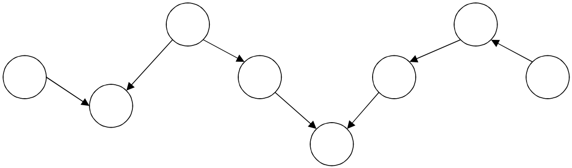
A path with 6 valves



From left to right

d-separation

A path with 6 valves

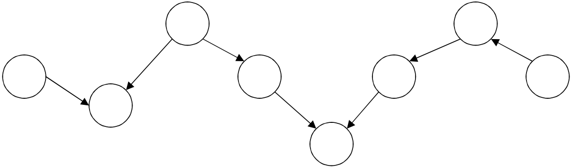


From left to right

convergent,

d-separation

A path with 6 valves

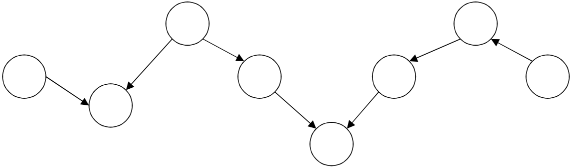


From left to right

convergent, divergent,

d-separation

A path with 6 valves

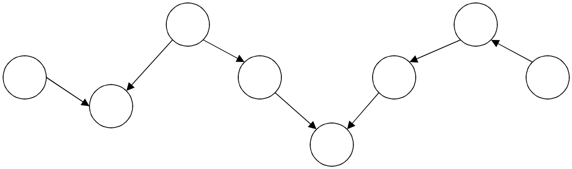


From left to right

convergent, divergent, sequential,

d-separation

A path with 6 valves

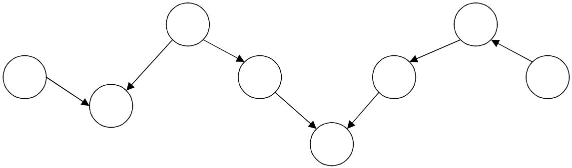


From left to right

convergent, divergent, sequential, convergent,

d-separation

A path with 6 valves

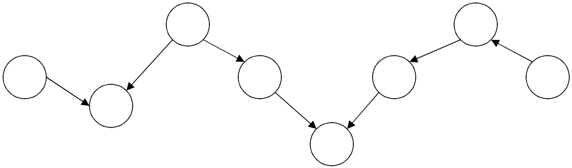


From left to right

convergent, divergent, sequential, convergent, sequential,

d-separation

A path with 6 valves



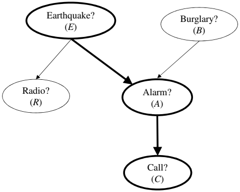
From left to right

convergent, divergent, sequential, convergent, sequential, and sequential.

d-separation

d-separation

sequential valve

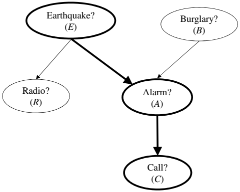


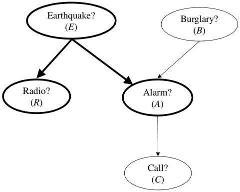
A is intermediary between cause E and effect C

d-separation

sequential valve

divergent valve

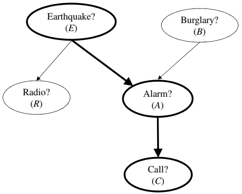
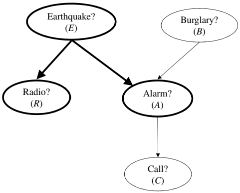
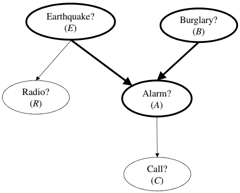




A is intermediary between cause E and effect C

E is common cause of effects R and A

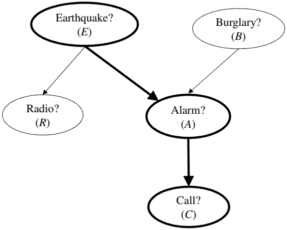
d-separation

sequential valve	divergent valve	convergent valve
		
A is intermediary between cause E and effect C	E is common cause of effects R and A	A is common effect of causes E and B

d-separation

Given that we know Z

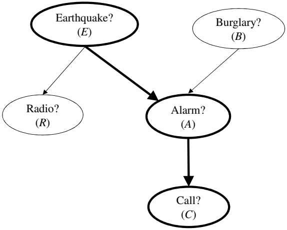
when is a sequential valve closed?



d-separation

Given that we know Z

when is a sequential valve closed?



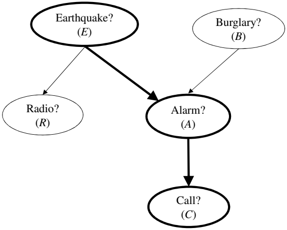
Valve $E \rightarrow A \rightarrow C$ is closed iff

we know the value of variable A, otherwise an earthquake E may change our belief in getting a call C.

d-separation

Given that we know $\mathbf{Z}$

when is a sequential valve closed?



Valve $E \rightarrow A \rightarrow C$ is closed iff

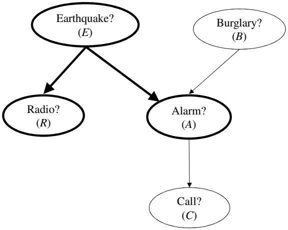
we know the value of variable A, otherwise an earthquake E may change our belief in getting a call C.

A sequential valve $\rightarrow W \rightarrow$ is closed iff variable W appears in $\mathbf{Z}$

d-separation

Given that we know Z

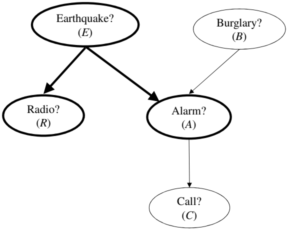
when is a divergent valve closed?



d-separation

Given that we know $\mathbf{Z}$

when is a divergent valve closed?



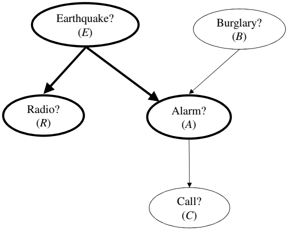
Valve $R \leftarrow E \rightarrow A$ is closed iff

we know the value of variable E, otherwise a radio report on an earthquake may change our belief in the alarm triggering.

d-separation

Given that we know $\mathbf{Z}$

when is a divergent valve closed?



Valve $R \leftarrow E \rightarrow A$ is closed iff

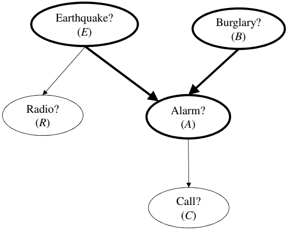
we know the value of variable E, otherwise a radio report on an earthquake may change our belief in the alarm triggering.

A divergent valve $\leftarrow W \rightarrow$ is closed iff variable W appears in $\mathbf{Z}$

d-separation

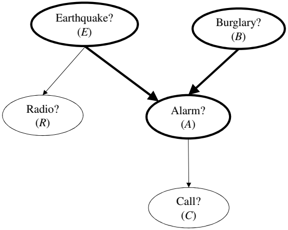
Given that we know **Z**

when is a convergent valve closed?



Given that we know $\mathbf{Z}$

when is a convergent valve closed?



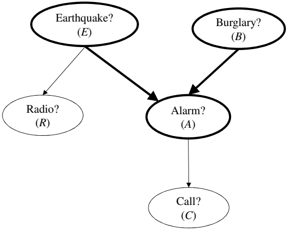
Valve $E \rightarrow A \leftarrow B$ is closed iff

neither the value of variable A nor the value of C are known, otherwise, a burglary may change our belief in an earthquake.

d-separation

Given that we know Z

when is a convergent valve closed?



Valve $E \rightarrow A \leftarrow B$ is closed iff

neither the value of variable A nor the value of C are known, otherwise, a burglary may change our belief in an earthquake.

A convergent valve $\rightarrow W \leftarrow$ is closed iff neither variable W nor any of its descendants appears in $\mathbf{Z}$

d-separation

$\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, iff

every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

d-separation

$\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, iff

every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

A path is blocked by $\mathbf{Z}$ iff

at least one valve on the path is closed given $\mathbf{Z}$

d-separation

$\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$, written $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, iff

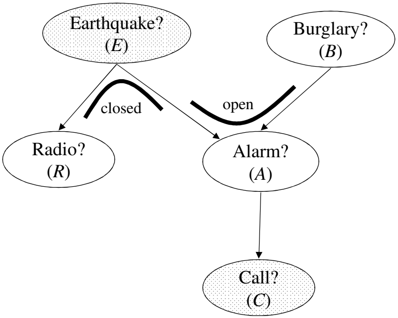
every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

A path is blocked by $\mathbf{Z}$ iff

at least one valve on the path is closed given $\mathbf{Z}$

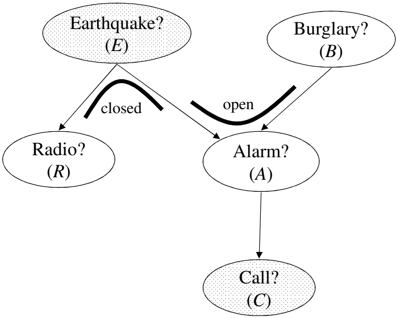
A path with no valves (i.e., $X \longrightarrow Y$) is never blocked.

d-separation



Are B and R d-separated by E and C?

d-separation

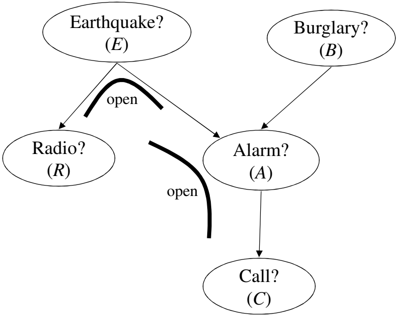


Are B and R d-separated by E and C?

Yes

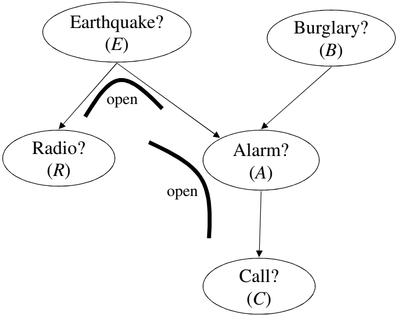
The closure of only one valve is sufficient to block the path, therefore, establishing d-separation.

d-separation



Are C and R d-separated?

d-separation

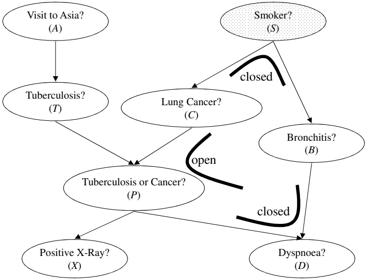


Are C and R d-separated?

No

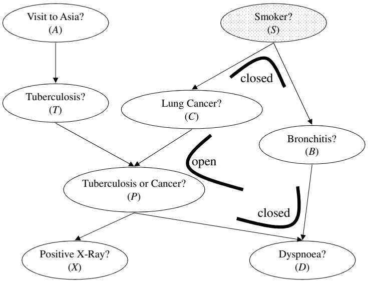
Both valves are open. Hence, the path is not blocked and d-separation does not hold.

d-separation



Are C and B d-separated by S?

d-separation



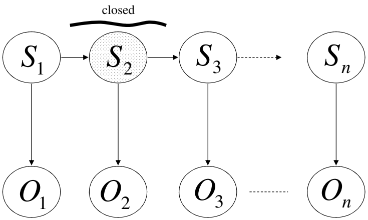
Are C and B d-separated by S?

Yes

Both paths between them are blocked by S.

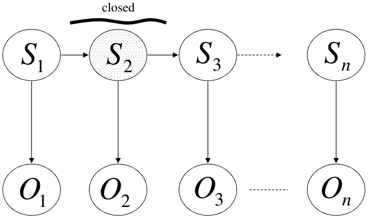
d-separation

Is $I_{\Pr}(S_1, S_2, \{S_3, S_4\})$ for any probability distribution $\Pr$ induced by the DAG?



d-separation

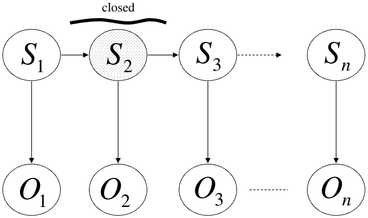
Is $I_{\mathrm{Pr}}(S_1, S_2, \{S_3, S_4\})$ for any probability distribution Pr induced by the DAG?



Valve $S_1 {\rightarrow} S_2 {\rightarrow} S_3$ on every path between S_1 and $\{S_3, S_4\}$

d-separation

Is $I_{\Pr}(S_1, S_2, \{S_3, S_4\})$ for any probability distribution $\Pr$ induced by the DAG?

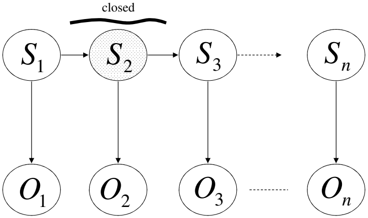


Valve $S_1 \rightarrow S_2 \rightarrow S_3$ on every path between S_1 and $\{S_3, S_4\}$

Valve is closed given S_2

d-separation

Is $I_{\Pr}(S_1, S_2, \{S_3, S_4\})$ for any probability distribution $\Pr$ induced by the DAG?



Valve $S_1 \rightarrow S_2 \rightarrow S_3$ on every path between S_1 and $\{S_3, S_4\}$

Valve is closed given S_2

Every path from S_1 to $\{S_3, S_4\}$ is blocked by S_2 and we have $\text{dsep}_G(S_1, S_2, \{S_3, S_4\})$

Blankets and Boundaries

A Markov blanket for variable X

is a set of variables which, when known, will render every other variable irrelevant to X

Blankets and Boundaries

A Markov blanket for variable X

is a set of variables which, when known, will render every other variable irrelevant to X

A Markov blanket $\mathbf{B}$ is minimal iff

no strict subset of $\mathbf{B}$ is also a Markov blanket.

Blankets and Boundaries

A Markov blanket for variable X

is a set of variables which, when known, will render every other variable irrelevant to X

A Markov blanket $\mathbf{B}$ is minimal iff

no strict subset of $\mathbf{B}$ is also a Markov blanket.

A minimal Markov blanket

is called a Markov Boundary.

Blankets and Boundaries

A Markov blanket for variable X

is a set of variables which, when known, will render every other variable irrelevant to X

A Markov blanket $\mathbf{B}$ is minimal iff

no strict subset of $\mathbf{B}$ is also a Markov blanket.

A minimal Markov blanket

is called a Markov Boundary.

The Markov Boundary is not unique

unless the distribution is strictly positive.

Blankets and Boundaries

If distribution $\Pr$ is induced by DAG G

then a Markov blanket for variable X with respect to $\Pr$ can be constructed using its parents, children, and spouses in DAG G

Blankets and Boundaries

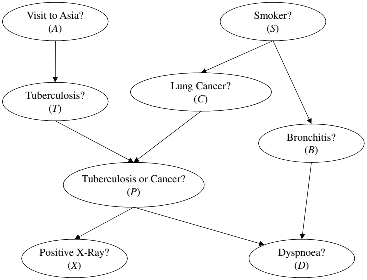
If distribution $\Pr$ is induced by DAG G

then a Markov blanket for variable X with respect to $\Pr$ can be constructed using its parents, children, and spouses in DAG G

Variable Y is a spouse of X iff

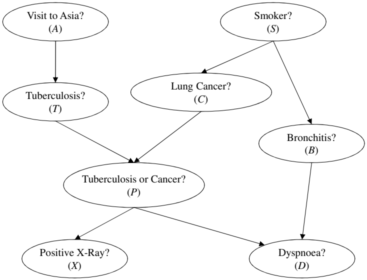
the two variables have a common child in DAG G

Blankets and Boundaries



Markov blanket for C

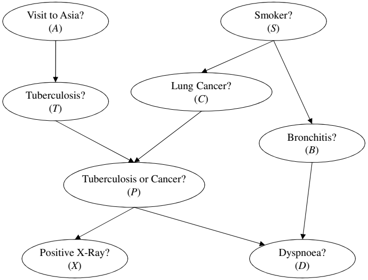
Blankets and Boundaries

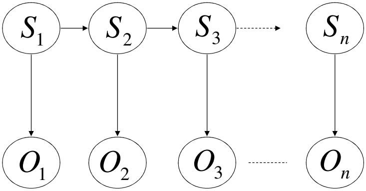


Markov blanket for C

S, P, T

Blankets and Boundaries



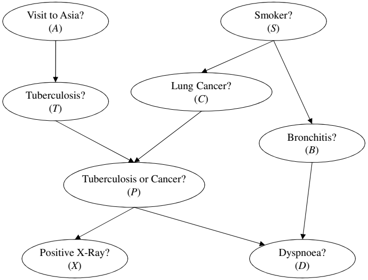


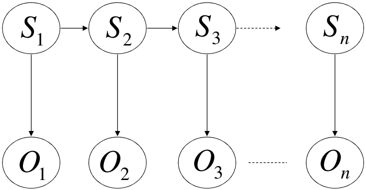
Markov blanket for C

S, P, T

Markov blanket for S_t, $t > 1$

Blankets and Boundaries





Markov blanket for S_t, $t > 1$

S_{t-1}, S_{t+1}, O_t

Markov blanket for C

S, P, T