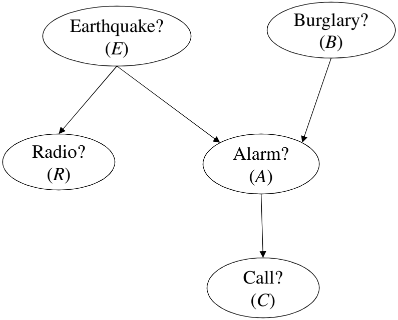
Chapter 4: Bayesian Networks

Adnan Darwiche[1]

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

Capturing Independence Graphically

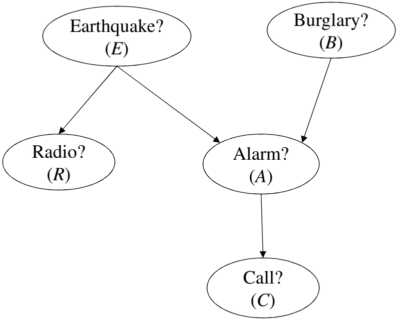


Assume that edges in this graph represent direct causal influences among these variables.

Example

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

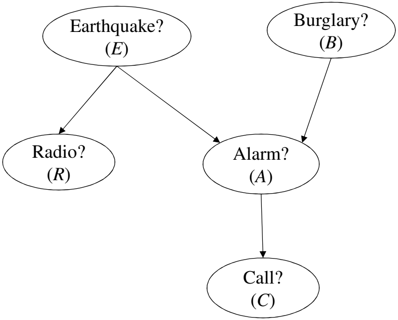
Capturing Independence Graphically



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.

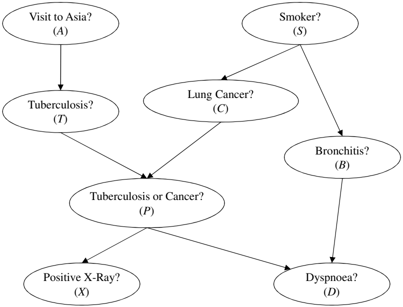
Capturing Independence Graphically



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.

$\text{Parents}(V)$

variables N with an edge from N to V

$\text{Descendants}(V)$

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

$\text{Non\_Descendants}(V)$

variables other than V, $\text{Parents}(V)$ and $\text{Descendants}(V)$

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

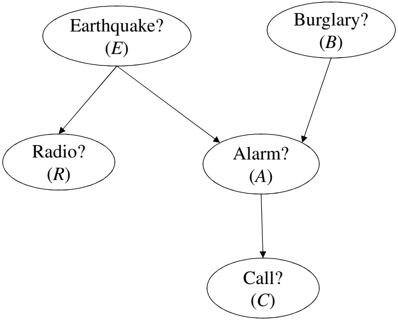
DAG as a causal structure

$\mathrm{Parents}(V)$ 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.

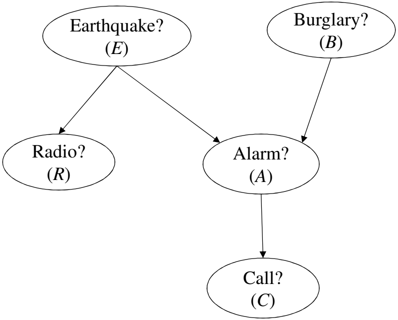
Capturing Independence Graphically



Markovian assumptions, Markov(G):

$I(C,$

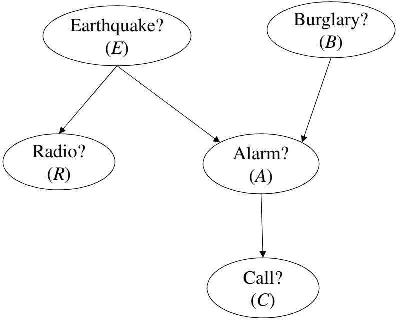
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

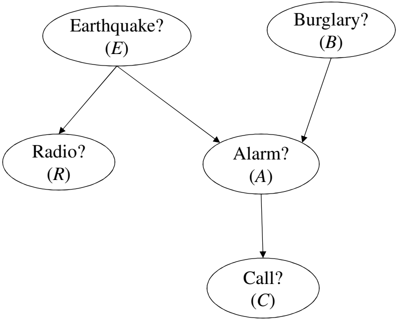
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

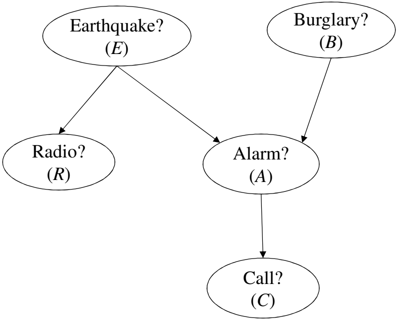
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

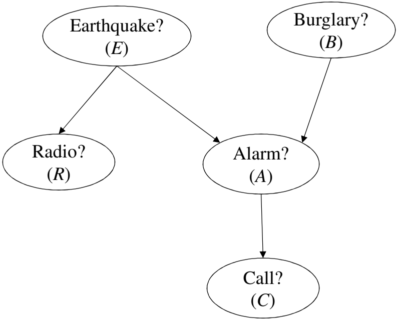
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

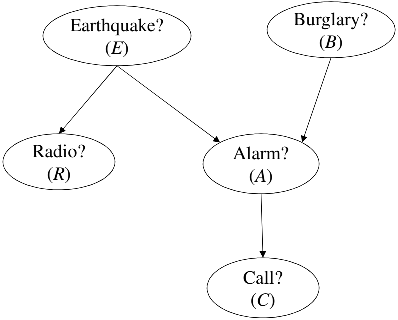
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

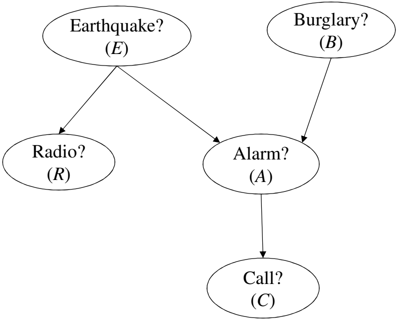
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

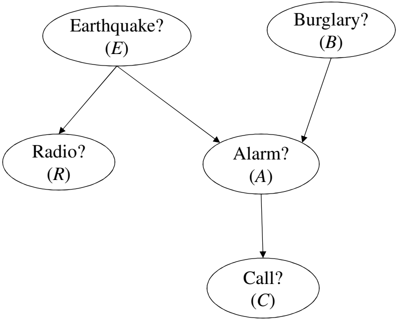
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

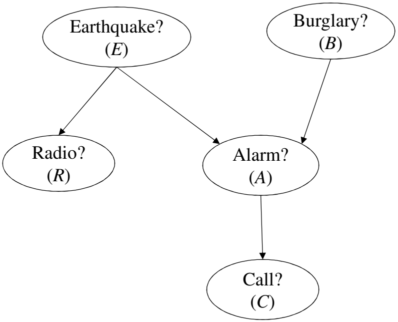
Capturing Independence Graphically



Markovian assumptions, Markov(G):

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

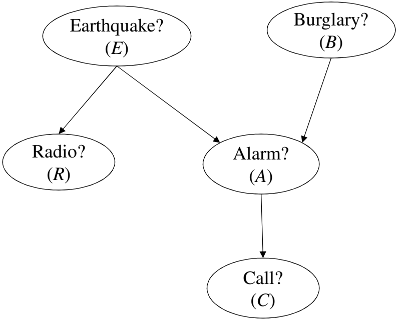
Capturing Independence Graphically



Markovian assumptions, 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



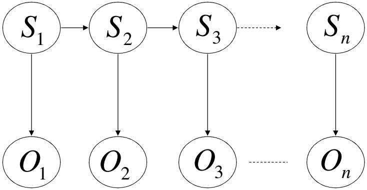
Markovian assumptions, Markov(G):

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

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

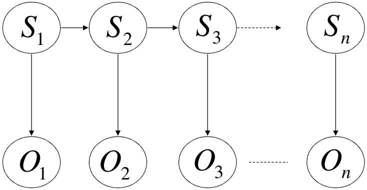
Capturing Independence Graphically

Hidden Markov Model



Capturing Independence Graphically

Hidden Markov Model

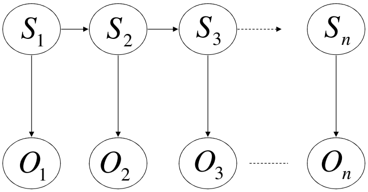


$S_1, S_2, \ldots, S_n$

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

Capturing Independence Graphically

Hidden Markov Model



$S_1, S_2, \ldots, S_n$

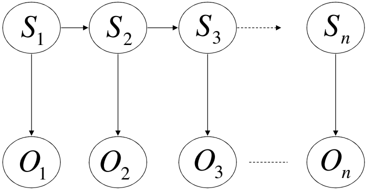
The state of a dynamic system at time points $1, 2, \ldots, n$

$O_1, O_2, \ldots, O_n$

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 $1, 2, \ldots, n$

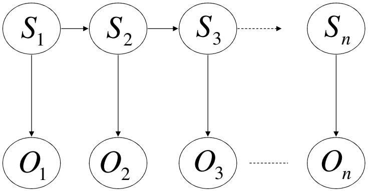
$O_1, O_2, \ldots, O_n$

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

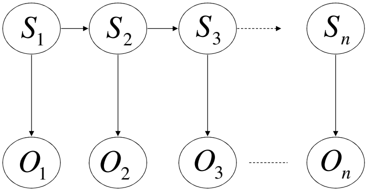


The Markovian assumptions imply that

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

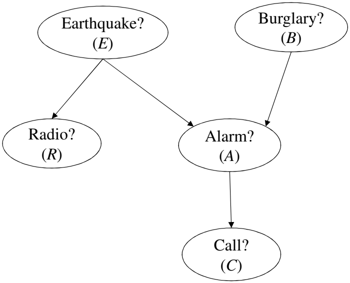
one tends to agree with the independencies declared by the DAG.

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

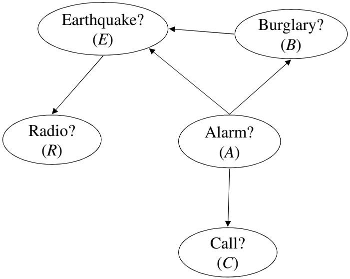
yet we agree with the independencies declared by the DAG.

Capturing Independence Graphically

DAG is causal

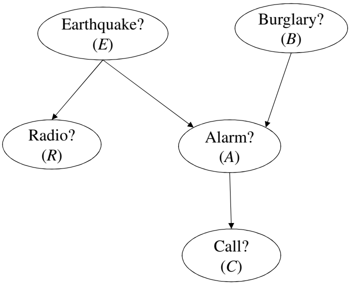


DAG is not causal

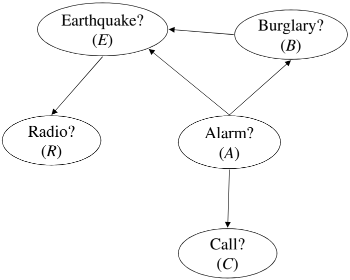


Capturing Independence Graphically

DAG is causal

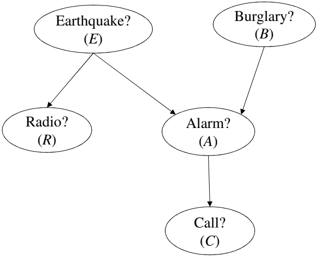


DAG is not causal



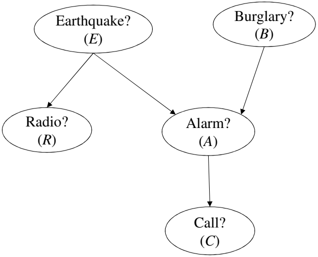
Every independence which is declared (or implied) by the second DAG is also declared (or implied) by the first one. Hence, if we accept the first DAG, then we must also accept the second.

Parameterizing the Independence Structure



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

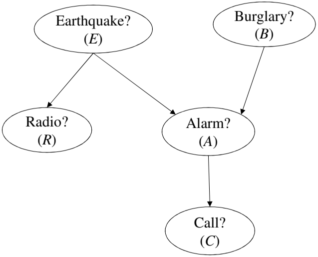
Parameterizing the Independence Structure



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

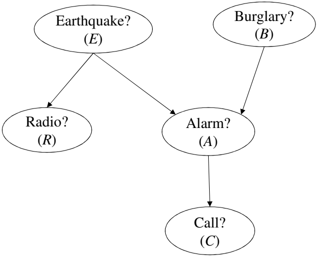


DAG G is a partial specification of our state of belief Pr

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



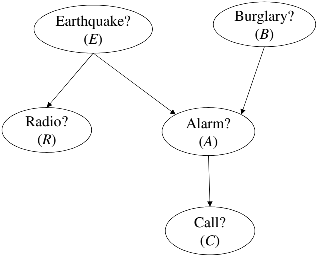
DAG G is a partial specification of our state of belief $\Pr$

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

This constrains $\Pr$ but does not uniquely define it.

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

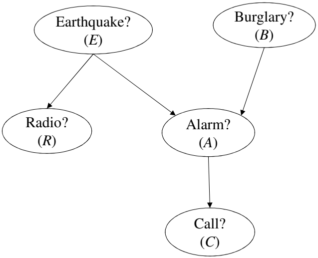
Parameterizing the Independence Structure



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



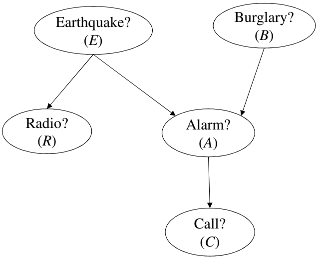
For every variable X and its parents $\mathbf{U}$

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

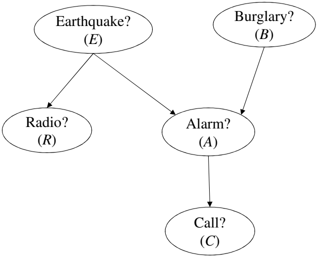


Conditional probabilities for variable C

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

Conditional Probability Table (CPT)

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

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.

Bayesian Networks

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

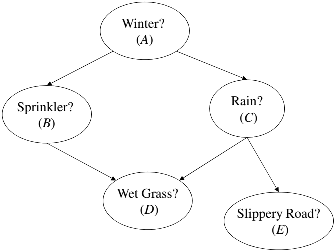
A Bayesian network for variables $\mathbf{Z}$ is a pair (G, Θ), where

- 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

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

The Distribution of a Bayesian Network

A Bayesian network induces distribution

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

The probability assigned to a network instantiation z

is the product of all network parameters that are compatible with **z**

The Distribution of a Bayesian Network

A Bayesian network induces distribution

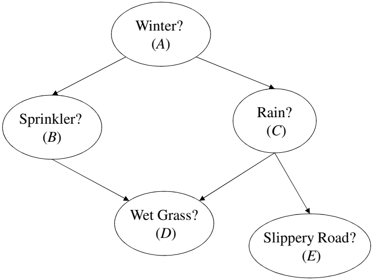
$$\Pr(\mathbf{z}) \stackrel{def}{=} \prod_{\theta_{x|\mathbf{u}} \sim \mathbf{z}} \theta_{x|\mathbf{u}}$$

The probability assigned to a network instantiation z

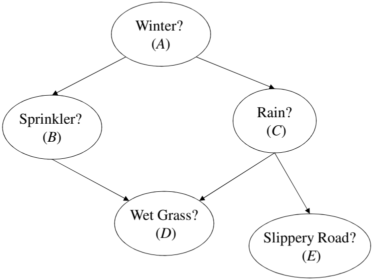
is the product of all network parameters that are compatible with $\mathbf{z}$

This is called the chain rule of Bayesian networks.

The Distribution of a Bayesian Network



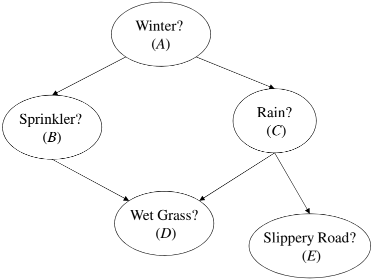
The Distribution of a Bayesian Network



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

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

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 $O(n \cdot d^{k+1})$

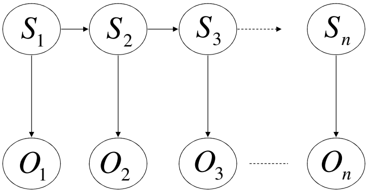
This number is quite reasonable

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

The Size of a Bayesian Network

Variable S_i has m values and similarly for variables O_i

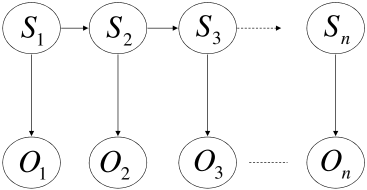
Hidden Markov Model



The Size of a Bayesian Network

Variable S_i has m values and similarly for variables O_i

Hidden Markov Model

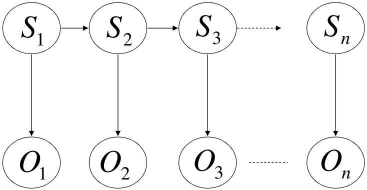


The CPT for any state variable S_i, $i > 1$, has m^2 parameters, known as transition probabilities.

The Size of a Bayesian Network

Variable S_i has m values and similarly for variables O_i

Hidden Markov Model



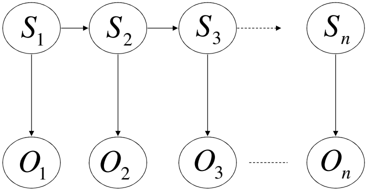
The CPT for any state variable S_i, $i > 1$, has m^2 parameters, known as transition probabilities.

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

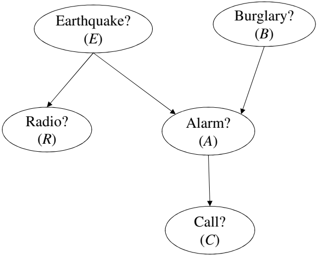


The CPT for any state variable S_i, $i > 1$, has m^2 parameters, known as transition probabilities.

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

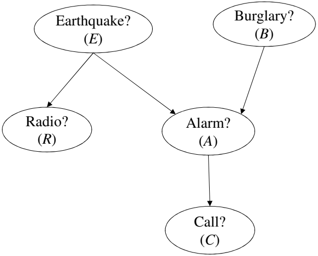
The CPT for S_1 has m parameters.

Properties of Probabilistic Independence



The distribution Pr specified by a Bayesian network (G, Θ) satisfies every independence assumption in $\mathrm{Markov}(G)$

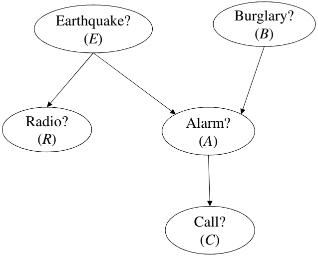
Properties of Probabilistic Independence



The distribution $\Pr$ specified by a Bayesian network (G, Θ) satisfies every independence assumption in $\mathrm{Markov}(G)$

These are not the only independencies satisfied by the distribution.

Properties of Probabilistic Independence



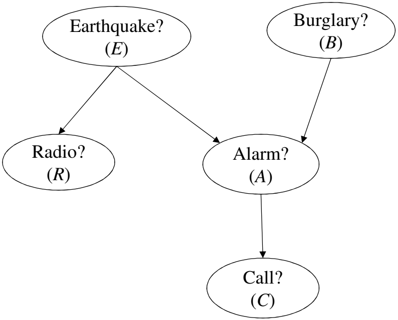
The distribution $\Pr$ specified by a Bayesian network (G, Θ) satisfies every independence assumption in $\mathrm{Markov}(G)$

These are not the only independencies satisfied by the distribution.

B and E independent given R

yet this independence is not part of $\mathrm{Markov}(G)$

Properties of Probabilistic Independence



B and E are independent given R

This independence and additional ones

follow from the ones in $\mathrm{Markov}(G)$ using a set of properties for probabilistic independence, known as the graphoid axioms.

Properties of Probabilistic Independence

Recall the definition of $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

Variables $\mathbf{X}$ independent of variables $\mathbf{Y}$ given variables $\mathbf{Z}$

Properties of Probabilistic Independence

Recall the definition of $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

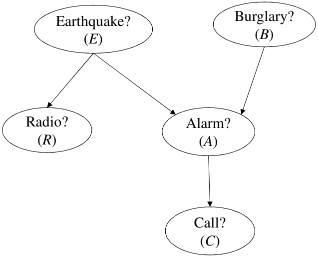
Variables **X** independent of variables **Y** given variables **Z**

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ iff

$$\mathrm{Pr}(\mathbf{x}|\mathbf{z}, \mathbf{y}) = \mathrm{Pr}(\mathbf{x}|\mathbf{z}) \quad \text{or } \mathrm{Pr}(\mathbf{y}, \mathbf{z}) = 0$$

for all instantiations $\mathbf{x}, \mathbf{y}, \mathbf{z}$

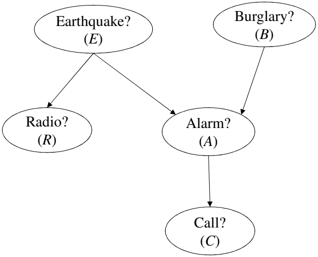
Symmetry



$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ iff $I_{\mathrm{Pr}}(\mathbf{Y}, \mathbf{Z}, \mathbf{X})$

Learning **y** does not influence our belief in **x** iff learning **x** does not influence our belief in **y**

Symmetry



$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ iff $I_{\mathrm{Pr}}(\mathbf{Y}, \mathbf{Z}, \mathbf{X})$

Learning **y** does not influence our belief in **x** iff learning **x** does not influence our belief in **y**

Example

From $\mathrm{Markov}(G)$, we have $I_{\mathrm{Pr}}(A, \{B, E\}, R)$. Using Symmetry, we get $I_{\mathrm{Pr}}(R, \{B, E\}, A)$ which is not part of $\mathrm{Markov}(G)$

Decomposition

$I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$

If learning **yw** does not influence our belief in **x**, then learning **y** alone, or learning **w** alone, will not influence our belief in **x** either.

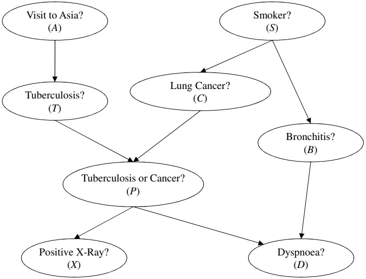
Decomposition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$

If learning **yw** does not influence our belief in **x**, then learning **y** alone, or learning **w** alone, will not influence our belief in **x** either.

If some information is irrelevant, then any part of it is also irrelevant.

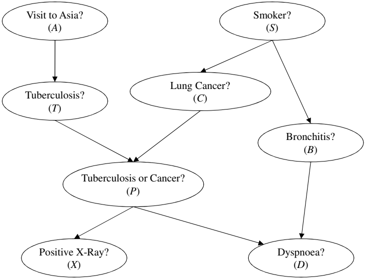
Decomposition



Example

$\text{Markov}(G)$ implies
$I(B, S, \{A, C, P, T, X\})$
Decomposition tells us $I(B, S, C)$

Decomposition



Example

$\text{Markov}(G)$ implies
$I(B, S, \{A, C, P, T, X\})$
Decomposition tells us $I(B, S, C)$

This independence holds in any probability distribution induced by a parametrization of DAG G. Yet, this independence is not part of the independencies declared by $\text{Markov}(G)$

Decomposition

An implication of Decomposition

$I_{\mathrm{Pr}}(X, \mathrm{Parents}(X), \mathbf{W})$ for every $\mathbf{W} \subseteq \mathrm{Non\_Descendants}(X)$

Decomposition

An implication of Decomposition

$I_{\mathrm{Pr}}(X, \mathrm{Parents}(X), \mathbf{W})$ for every $\mathbf{W} \subseteq \mathrm{Non\_Descendants}(X)$

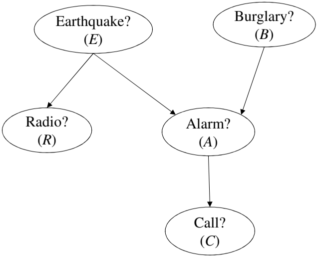
Every variable X is conditionally independent of any subset of its non-descendants given its parents.

This is a strengthening of the independence statements declared by $\mathrm{Markov}(G)$, which is a special case when $\mathbf{W}$ contains all non-descendants of X

Decomposition: The chain rule for Bayesian networks

By the chain rule of probability calculus

$$\Pr(r, c, a, e, b) = \Pr(r|c, a, e, b)\Pr(c|a, e, b)\Pr(a|e, b)\Pr(e|b)\Pr(b)$$



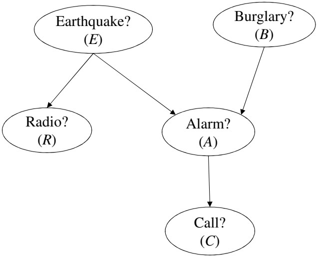
By Decomposition

$$\begin{aligned} \Pr(r|c, a, e, b) &= \Pr(r|e) \\ \Pr(c|a, e, b) &= \Pr(c|a) \\ \Pr(e|b) &= \Pr(e) \end{aligned}$$

Decomposition: The chain rule for Bayesian networks

By the chain rule of probability calculus

$$\Pr(r, c, a, e, b) = \Pr(r|c, a, e, b)\Pr(c|a, e, b)\Pr(a|e, b)\Pr(e|b)\Pr(b)$$



By Decomposition

$$\begin{aligned} \Pr(r|c, a, e, b) &= \Pr(r|e) \\ \Pr(c|a, e, b) &= \Pr(c|a) \\ \Pr(e|b) &= \Pr(e) \end{aligned}$$

This leads to the chain rule of Bayesian networks

$$\begin{aligned} \Pr(r, c, a, e, b) &= \Pr(r|e)\Pr(c|a)\Pr(a|e, b)\Pr(e)\Pr(b) \\ &= \theta_{r|e}\ \theta_{c|a}\ \theta_{a|e,b}\ \theta_e\ \theta_b \end{aligned}$$

Decomposition

Proof generalizes to any Bayesian network over variables $\mathbf{Z}$

as long as we order variables $\mathbf{Z}$ such that the parents $\mathbf{U}$ of each variable X appear after X in the order.

Decomposition

Proof generalizes to any Bayesian network over variables $\mathbf{Z}$

as long as we order variables $\mathbf{Z}$ such that the parents $\mathbf{U}$ of each variable X appear after X in the order.

This ordering constraint ensures two things:

- For every term $\Pr(x|\alpha)$ that results from applying the chain rule to $\Pr(\mathbf{z})$, some instantiation $\mathbf{u}$ of parents $\mathbf{U}$ is guaranteed to be in α.

Decomposition

Proof generalizes to any Bayesian network over variables $\mathbf{Z}$

as long as we order variables $\mathbf{Z}$ such that the parents $\mathbf{U}$ of each variable X appear after X in the order.

This ordering constraint ensures two things:

- For every term $\Pr(x|\alpha)$ that results from applying the chain rule to $\Pr(\mathbf{z})$, some instantiation $\mathbf{u}$ of parents $\mathbf{U}$ is guaranteed to be in α.
- The only other variables appearing in α, beyond parents $\mathbf{U}$, must be non-descendants of X.

Decomposition

Proof generalizes to any Bayesian network over variables $\mathbf{Z}$

as long as we order variables $\mathbf{Z}$ such that the parents $\mathbf{U}$ of each variable X appear after X in the order.

This ordering constraint ensures two things:

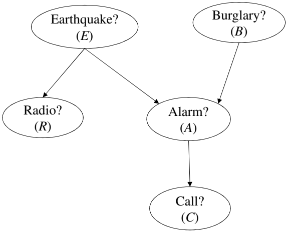
- For every term $\Pr(x|\alpha)$ that results from applying the chain rule to $\Pr(\mathbf{z})$, some instantiation $\mathbf{u}$ of parents $\mathbf{U}$ is guaranteed to be in α.
- The only other variables appearing in α, beyond parents $\mathbf{U}$, must be non-descendants of X.

Hence, the term $\Pr(x|\alpha)$ must equal the network parameter $\theta_{x|\mathbf{u}}$ by Decomposition.

Decomposition

The variable ordering c, a, r, b, e gives

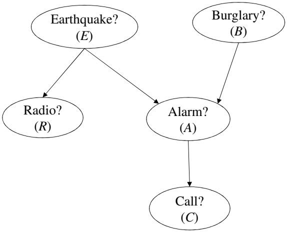
$$\Pr(c, a, r, b, e) = \Pr(c|a, r, b, e)\Pr(a|r, b, e)\Pr(r|b, e)\Pr(b|e)\Pr(e)$$



Decomposition

The variable ordering c, a, r, b, e gives

$$\Pr(c, a, r, b, e) = \Pr(c|a, r, b, e)\Pr(a|r, b, e)\Pr(r|b, e)\Pr(b|e)\Pr(e)$$



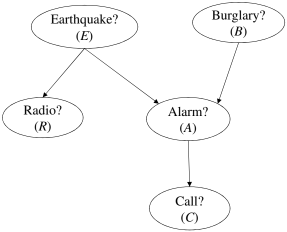
By Decomposition

$$\begin{aligned} \Pr(c|a, r, b, e) &= \Pr(c|a) \\ \Pr(a|r, b, e) &= \Pr(a|b, e) \\ \Pr(r|b, e) &= \Pr(r|e) \\ \Pr(b|e) &= \Pr(b) \end{aligned}$$

Decomposition

The variable ordering c, a, r, b, e gives

$$\Pr(c, a, r, b, e) = \Pr(c|a, r, b, e)\Pr(a|r, b, e)\Pr(r|b, e)\Pr(b|e)\Pr(e)$$



By Decomposition

$$\begin{aligned} \Pr(c|a, r, b, e) &= \Pr(c|a) \\ \Pr(a|r, b, e) &= \Pr(a|b, e) \\ \Pr(r|b, e) &= \Pr(r|e) \\ \Pr(b|e) &= \Pr(b) \end{aligned}$$

We then have

$$\begin{aligned} \Pr(c, a, r, b, e) &= \Pr(c|a)\Pr(a|b, e)\Pr(r|e)\Pr(b)\Pr(e) \\ &= \theta_{c|a}\ \theta_{a|b,e}\ \theta_{r|e}\ \theta_b\ \theta_e \end{aligned}$$

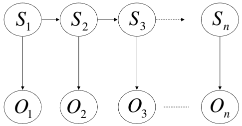
Decomposition

The variable ordering $o_n, \ldots, o_1, s_n, \ldots, s_1$ gives

$$\Pr(o_n, \ldots, o_1, s_n, \ldots, s_1) =$$
$$\Pr(o_n|o_{n-1} \ldots, o_1, s_n, \ldots, s_1) \ldots \Pr(o_1|s_n, \ldots, s_1)\Pr(s_n|s_{n-1} \ldots, s_1) \ldots \Pr(s_1)$$

By Decomposition

$$\begin{aligned}\Pr(o_n, \ldots, o_1, s_n, \ldots, s_1) \\ &= \Pr(o_n|s_n) \ldots \Pr(o_1|s_1)\Pr(s_n|s_{n-1}) \ldots \Pr(s_1) \\ &= \theta_{o_n|s_n} \ldots \theta_{o_1|s_1}\ \theta_{s_n|s_{n-1}} \ldots \theta_{s_1}.\end{aligned}$$



$\Pr(o_n, \ldots, o_1, s_n, \ldots, s_1)$

is now expressed as a product of network parameters.

Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Composition is the opposite of Decomposition.

Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Composition is the opposite of Decomposition.

Composition does not hold in general

Two pieces of information may each be irrelevant on their own, yet their combination may be relevant.

Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Composition is the opposite of Decomposition.

Composition does not hold in general

Two pieces of information may each be irrelevant on their own, yet their combination may be relevant.

Example

An exclusive-or gate with uniform distribution on each input. Each input on its own is irrelevant to the output. Yet, both inputs together are relevant.

Weak Union

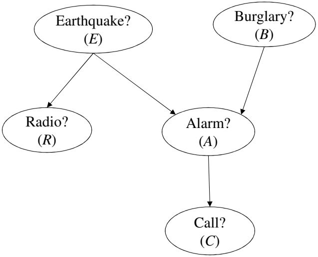
$I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$

If the information **yw** is not relevant to our belief in **x**, then the partial information **y** will not make the rest of the information, **w**, relevant.

Weak Union

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$

If the information **yw** is not relevant to our belief in **x**, then the partial information **y** will not make the rest of the information, **w**, relevant.



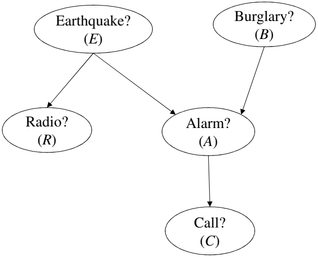
Markov(G) gives

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

Weak Union

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$

If the information **yw** is not relevant to our belief in **x**, then the partial information **y** will not make the rest of the information, **w**, relevant.



$\mathrm{Markov}(G)$ gives

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

By Weak Union

$I(C, \{A, B, E\}, R)$ which is not part of $\mathrm{Markov}(G)$

Weak Union

An implication of Weak Union

$$I_{\mathrm{Pr}}(X, \mathrm{Parents}(X) \cup \mathbf{W}, \mathrm{Non\_Descendants}(X) \setminus \mathbf{W})$$

for any $\mathbf{W} \subseteq \mathrm{Non\_Descendants}(X)$

Weak Union

An implication of Weak Union

$$I_{\mathrm{Pr}}(X, \mathrm{Parents}(X) \cup \mathbf{W}, \mathrm{Non\_Descendants}(X) \setminus \mathbf{W})$$

for any $\mathbf{W} \subseteq \mathrm{Non\_Descendants}(X)$

Each variable X in DAG G is independent of any of its non-descendants given its parents and the remaining non-descendants.

Weak Union

An implication of Weak Union

$$I_{\mathrm{Pr}}(X, \mathrm{Parents}(X) \cup \mathbf{W}, \mathrm{Non\_Descendants}(X) \setminus \mathbf{W})$$

for any $\mathbf{W} \subseteq \mathrm{Non\_Descendants}(X)$

Each variable X in DAG G is independent of any of its non-descendants given its parents and the remaining non-descendants.

This can be viewed as a strengthening of the independencies declared by $\mathrm{Markov}(G)$, which fall as a special case when the set $\mathbf{W}$ is empty.

Contraction

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If after learning the irrelevant information **y**, the information **w** is found to be irrelevant to our belief in **x**, then the combined information **yw** must have been irrelevant from the beginning.

Contraction

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If after learning the irrelevant information **y**, the information **w** is found to be irrelevant to our belief in **x**, then the combined information **yw** must have been irrelevant from the beginning.

Compare Contraction with Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Contraction

$I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If after learning the irrelevant information **y**, the information **w** is found to be irrelevant to our belief in **x**, then the combined information **yw** must have been irrelevant from the beginning.

Compare Contraction with Composition

$I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

One can view Contraction as a weaker version of Composition. Recall that Composition does not hold for probability distributions.

Strictly Positive Distributions

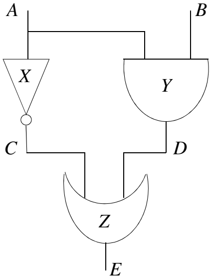
A strictly positive distribution

assign a non-zero probability to every consistent event.

Strictly Positive Distributions

A strictly positive distribution

assign a non-zero probability to every consistent event.



Example

A strictly positive distribution cannot represent the behavior of Inverter X as it will have to assign the probability zero to the event $A=$true, $C=$true

A strictly positive distribution cannot capture logical constraints.

Intersection

Holds only for strictly positive distributions.

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

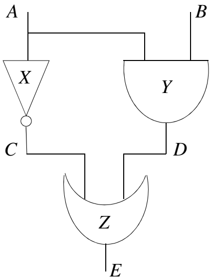
If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.

Intersection

Holds only for strictly positive distributions.

$I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.

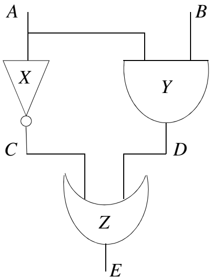


Intersection

Holds only for strictly positive distributions.

$I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.



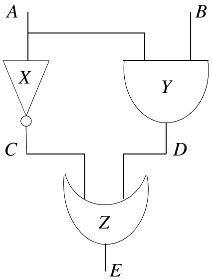
Given A, C irrelevant to E

Intersection

Holds only for strictly positive distributions.

$I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.



Given A, C irrelevant to E

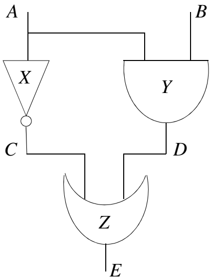
Given C, A irrelevant to E

Intersection

Holds only for strictly positive distributions.

$I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

If information **w** is irrelevant given **y**, and **y** is irrelevant given **w**, then combined information **yw** is irrelevant to start with.



Given A, C irrelevant to E

Given C, A irrelevant to E

Yet

A and C are not irrelevant to E

The Graphoid Axioms

Triviality: $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \emptyset)$

The Graphoid Axioms

Triviality: $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \emptyset)$

Symmetry, Decomposition, Weak Union, and Contraction, combined with Triviality, are known as the graphoid axioms.

The Graphoid Axioms

Triviality: $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \emptyset)$

Symmetry, Decomposition, Weak Union, and Contraction, combined with Triviality, are known as the graphoid axioms.

With Intersection, the set is known as the positive graphoid axioms.

The Graphoid Axioms

Triviality: $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \emptyset)$

Symmetry, Decomposition, Weak Union, and Contraction, combined with Triviality, are known as the graphoid axioms.

With Intersection, the set is known as the positive graphoid axioms.

Decomposition, Weak Union, and Contraction in one statement

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ iff $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$

The Graphoid Axioms

Triviality: $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \emptyset)$

Symmetry, Decomposition, Weak Union, and Contraction, combined with Triviality, are known as the graphoid axioms.

With Intersection, the set is known as the positive graphoid axioms.

Decomposition, Weak Union, and Contraction in one statement

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$ iff $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$

The terms semi-graphoid and graphoid are sometimes used instead of graphoid and positive graphoid, respectively.

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.

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

iff every path between a node in **X** and a node in **Y** is blocked by **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.

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

iff every path between a node in **X** and a node in **Y** is blocked by **Z**

The definition of d-separation relies on

the notion of blocking a path by a set of variables **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.

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

iff every path between a node in **X** and a node in **Y** is blocked by **Z**

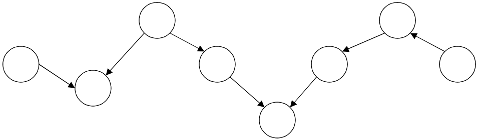
The definition of d-separation relies on

the notion of blocking a path by a set of variables **Z**

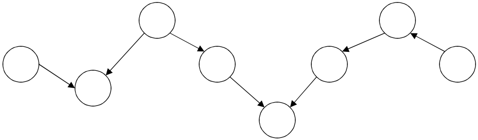
$\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ implies $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

for every probability distribution $\Pr$ induced by G

d-separation



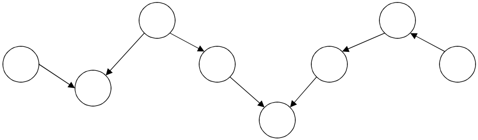
d-separation



View the path as a pipe

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

d-separation



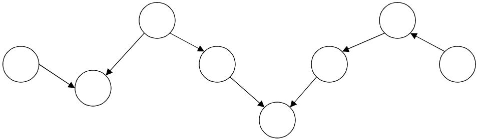
View the path as a pipe

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 $\rightarrow W \rightarrow$

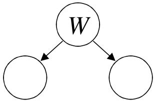
d-separation

The type of a valve

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

sequential $\rightarrow W \rightarrow$

divergent $\leftarrow W \rightarrow$



d-separation

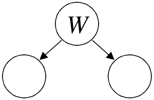
The type of a valve

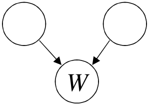
is determined by its relationship to its neighbors on the path.

sequential $\rightarrow W \rightarrow$

divergent $\leftarrow W \rightarrow$

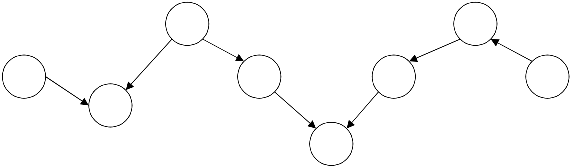
convergent $\rightarrow W \leftarrow$





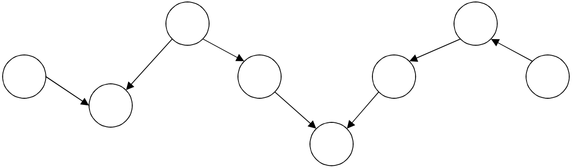
d-separation

A path with 6 valves



d-separation

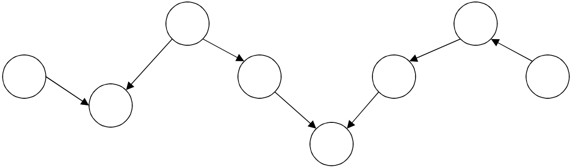
A path with 6 valves



From left to right

d-separation

A path with 6 valves

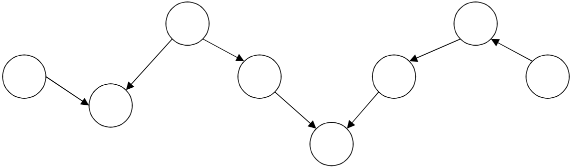


From left to right

convergent,

d-separation

A path with 6 valves

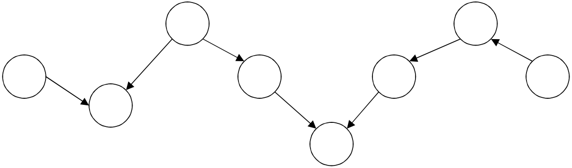


From left to right

convergent, divergent,

d-separation

A path with 6 valves

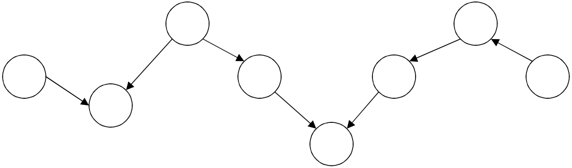


From left to right

convergent, divergent, sequential,

d-separation

A path with 6 valves

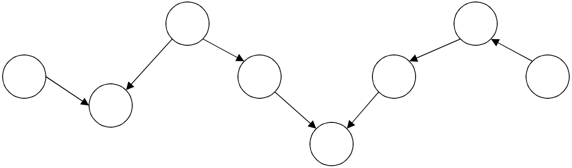


From left to right

convergent, divergent, sequential, convergent,

d-separation

A path with 6 valves

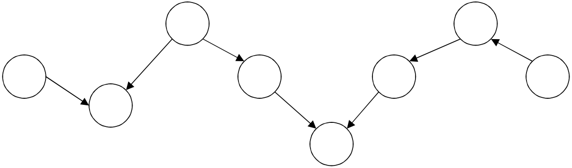


From left to right

convergent, divergent, sequential, convergent, sequential,

d-separation

A path with 6 valves



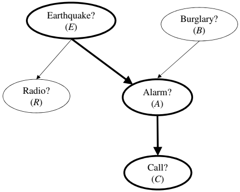
From left to right

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

d-separation

d-separation

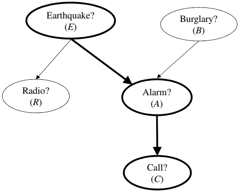
sequential valve



A is intermediary between cause E and effect C

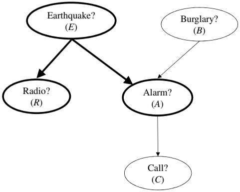
d-separation

sequential valve



A is intermediary between cause *E* and effect *C*

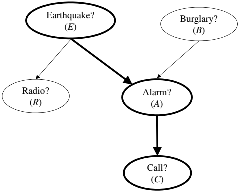
divergent valve



E is common cause of effects *R* and *A*

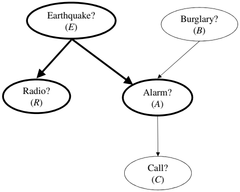
d-separation

sequential valve



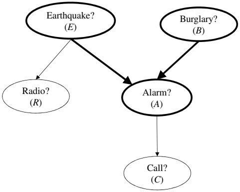
A is intermediary between cause E and effect C

divergent valve



E is common cause of effects R and A

convergent valve

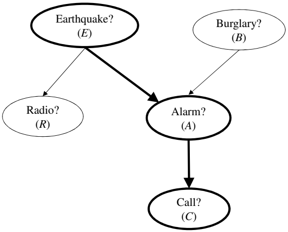


A is common effect of causes E and B

d-separation

Given that we know Z

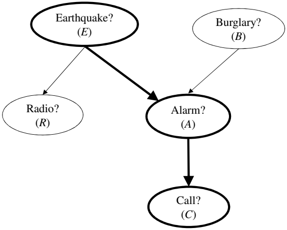
when is a sequential valve closed?



d-separation

Given that we know $\mathbf{Z}$

when is a sequential valve closed?



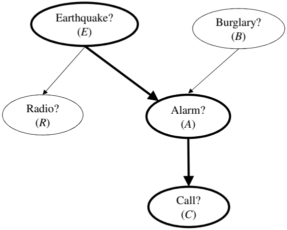
Valve $E \rightarrow A \rightarrow C$ is closed iff

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

d-separation

Given that we know Z

when is a sequential valve closed?



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

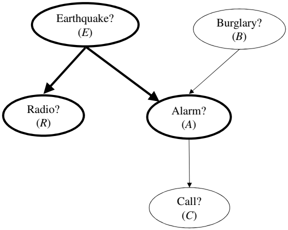
we know the value of variable A, otherwise an earthquake E may change our belief in getting a call C.

A sequential valve $\rightarrow W \rightarrow$ is closed iff variable W appears in **Z**

d-separation

Given that we know **Z**

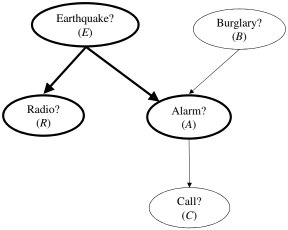
when is a divergent valve closed?



d-separation

Given that we know Z

when is a divergent valve closed?



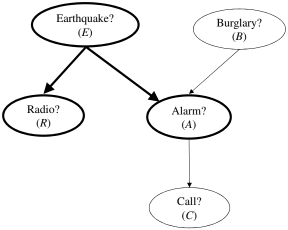
Valve $R \leftarrow E \rightarrow A$ is closed iff

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

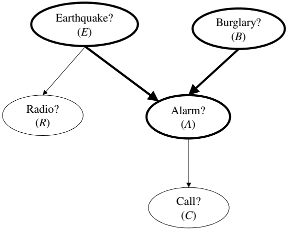
we know the value of variable E, otherwise a radio report on an earthquake may change our belief in the alarm triggering.

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

d-separation

Given that we know **Z**

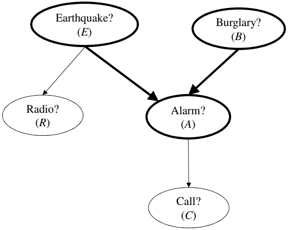
when is a convergent valve closed?



d-separation

Given that we know $\mathbf{Z}$

when is a convergent valve closed?



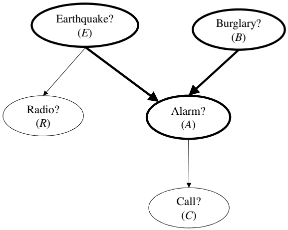
Valve $E \rightarrow A \leftarrow B$ is closed iff

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 $\mathbf{Z}$

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.

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 $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, iff

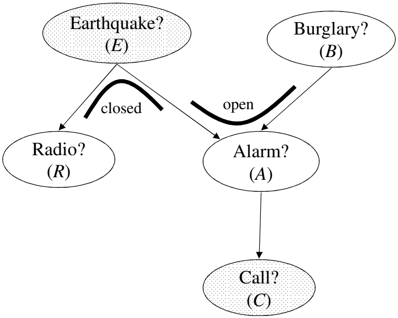
every path between a node in $\mathbf{X}$ and a node in $\mathbf{Y}$ is blocked by $\mathbf{Z}$

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

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

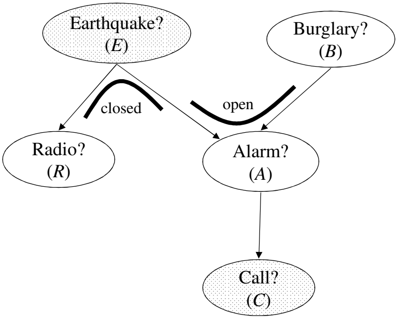
A path with no valves (i.e., $X \rightarrow Y$) is never blocked.

d-separation



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

d-separation

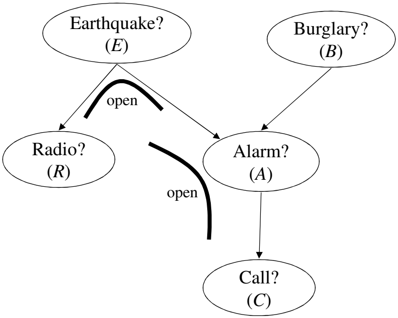


Are B and R d-separated by E and C?

Yes

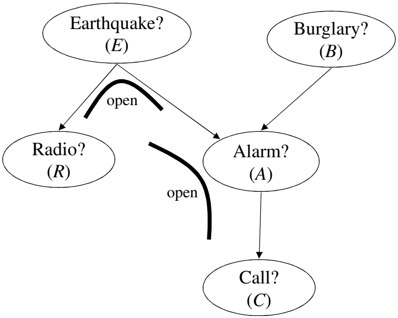
The closure of only one valve is sufficient to block the path, therefore, establishing d-separation.

d-separation



Are C and R d-separated?

d-separation

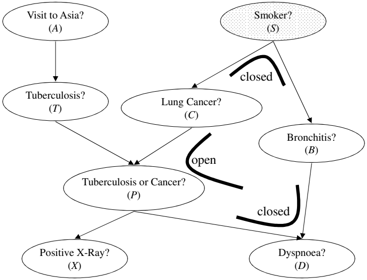


Are C and R d-separated?

No

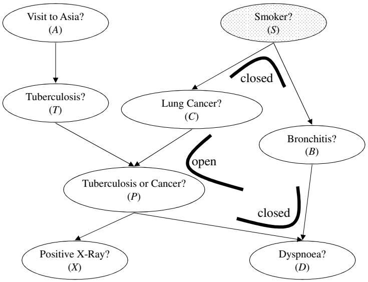
Both valves are open. Hence, the path is not blocked and d-separation does not hold.

d-separation



Are C and B d-separated by S?

d-separation



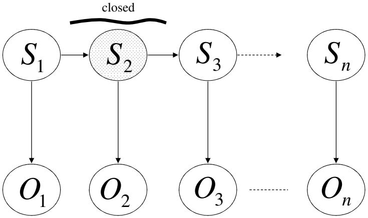
Are C and B d-separated by S?

Yes

Both paths between them are blocked by S.

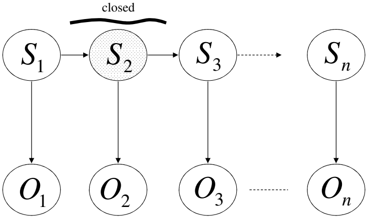
d-separation

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



d-separation

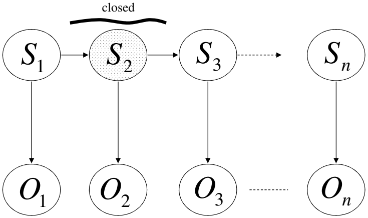
Is $I_{\Pr}(S_1, S_2, \{S_3, S_4\})$ for any probability distribution $\Pr$ induced by the DAG?



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

d-separation

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

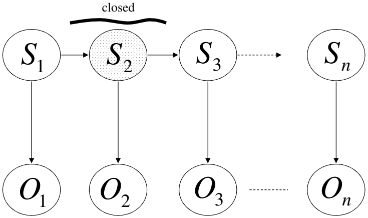


Valve $S_1 \rightarrow S_2 \rightarrow S_3$ on every path between S_1 and $\{S_3, S_4\}$

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 $\text{dsep}_G(S_1, S_2, \{S_3, S_4\})$

Complexity of d-separation

The definition of d-separation, $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, calls for

considering all paths connecting a node in $\mathbf{X}$ with a node in $\mathbf{Y}$. The number of such paths can be exponential, yet one can implement the test without having to enumerate these paths explicitly.

Complexity of d-separation

The definition of d-separation, $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, calls for

considering all paths connecting a node in $\mathbf{X}$ with a node in $\mathbf{Y}$. The number of such paths can be exponential, yet one can implement the test without having to enumerate these paths explicitly.

Deciding $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is equivalent to testing whether $\mathbf{X}$ and $\mathbf{Y}$ are disconnected in a new DAG G' obtained by pruning DAG G

Complexity of d-separation

The definition of d-separation, $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, calls for

considering all paths connecting a node in $\mathbf{X}$ with a node in $\mathbf{Y}$. The number of such paths can be exponential, yet one can implement the test without having to enumerate these paths explicitly.

Deciding $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is equivalent to testing whether $\mathbf{X}$ and $\mathbf{Y}$ are disconnected in a new DAG G' obtained by pruning DAG G

- Delete any leaf node W from DAG G as long as W not in $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$. Repeat until no more nodes can be deleted.

Complexity of d-separation

The definition of d-separation, $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, calls for

considering all paths connecting a node in $\mathbf{X}$ with a node in $\mathbf{Y}$. The number of such paths can be exponential, yet one can implement the test without having to enumerate these paths explicitly.

Deciding $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is equivalent to testing whether $\mathbf{X}$ and $\mathbf{Y}$ are disconnected in a new DAG G' obtained by pruning DAG G

- Delete any leaf node W from DAG G as long as W not in $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$. Repeat until no more nodes can be deleted.
- Delete all edges outgoing from nodes in $\mathbf{Z}$.

Complexity of d-separation

The definition of d-separation, $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, calls for

considering all paths connecting a node in $\mathbf{X}$ with a node in $\mathbf{Y}$. The number of such paths can be exponential, yet one can implement the test without having to enumerate these paths explicitly.

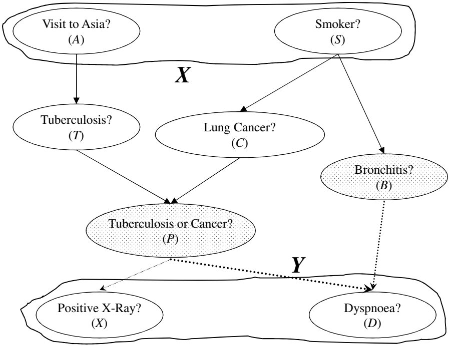
Deciding $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is equivalent to testing whether $\mathbf{X}$ and $\mathbf{Y}$ are disconnected in a new DAG G' obtained by pruning DAG G

- Delete any leaf node W from DAG G as long as W not in $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$. Repeat until no more nodes can be deleted.
- Delete all edges outgoing from nodes in $\mathbf{Z}$.

Decided in time and space that are linear in the size of DAG G

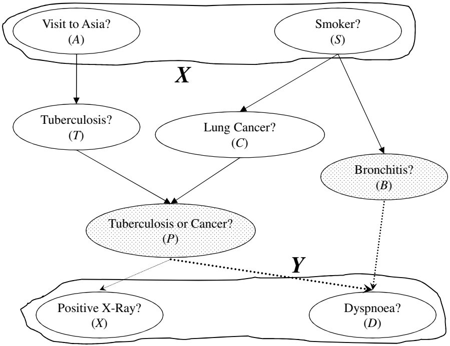
Complexity of d-separation

Nodes in **Z** are shaded. Pruned nodes and edges are dotted.



Complexity of d-separation

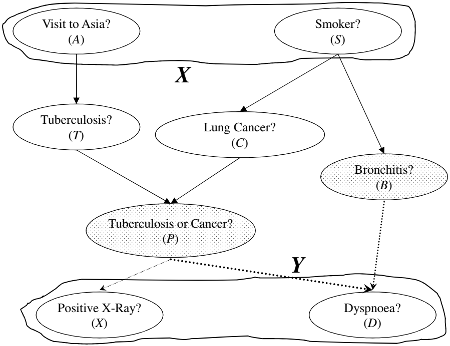
Nodes in **Z** are shaded. Pruned nodes and edges are dotted.



Is $\mathbf{X} = \{A, S\}$ d-separated from $\mathbf{Y} = \{D, X\}$ by $\mathbf{Z} = \{B, P\}$?

Complexity of d-separation

Nodes in **Z** are shaded. Pruned nodes and edges are dotted.

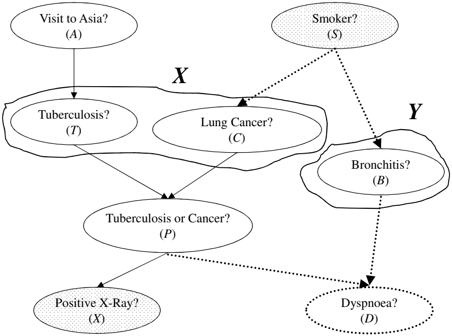


Is $\mathbf{X} = \{A, S\}$ d-separated from $\mathbf{Y} = \{D, X\}$ by $\mathbf{Z} = \{B, P\}$?

Yes

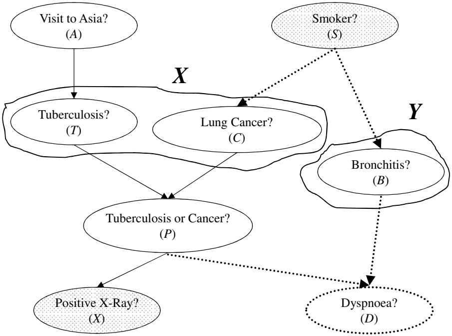
Complexity of d-separation

Nodes in **Z** are shaded. Pruned nodes and edges are dotted.



Complexity of d-separation

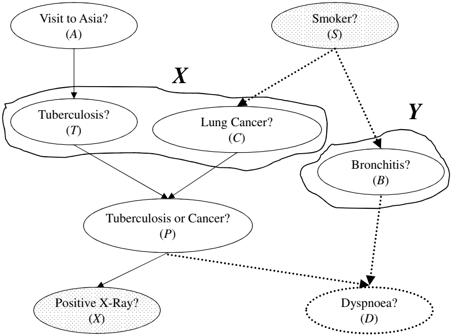
Nodes in **Z** are shaded. Pruned nodes and edges are dotted.



Is $\mathbf{X} = \{T, C\}$ d-separated from $\mathbf{Y} = \{B\}$ by $\mathbf{Z} = \{S, X\}$?

Complexity of d-separation

Nodes in **Z** are shaded. Pruned nodes and edges are dotted.



Is $\mathbf{X} = \{T, C\}$ d-separated from $\mathbf{Y} = \{B\}$ by $\mathbf{Z} = \{S, X\}$?

Yes

Soundness of d-separation

The d-separation test is sound

If distribution Pr is induced by Bayesian network (G, Θ), then

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

Soundness of d-separation

The d-separation test is sound

If distribution $\Pr$ is induced by Bayesian network (G, Θ), then

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

The proof of soundness is constructive

showing that every independence claimed by d-separation can indeed be derived using the graphoid axioms.

Completeness of d-separation

d-separation is not complete

Completeness of d-separation

d-separation is not complete

- Consider a network with three binary variables $X \longrightarrow Y \longrightarrow Z$

Completeness of d-separation

d-separation is not complete

- Consider a network with three binary variables $X \rightarrow Y \rightarrow Z$
- Z is not d-separated from X

Completeness of d-separation

d-separation is not complete

- Consider a network with three binary variables $X \rightarrow Y \rightarrow Z$
- Z is not d-separated from X
- Z can be independent of X in a probability distribution induced by this network.

Completeness of d-separation

d-separation is not complete

- Consider a network with three binary variables $X \rightarrow Y \rightarrow Z$
- Z is not d-separated from X
- Z can be independent of X in a probability distribution induced by this network.

Example

Choose the CPT for variable Y so that $\theta_{y|x} = \theta_{y|\bar{x}}$

Y independent of X since

- $\Pr(y) = \Pr(y|x) = \Pr(y|\bar{x})$ and
- $\Pr(\bar{y}) = \Pr(\bar{y}|x) = \Pr(\bar{y}|\bar{x})$

Z is also independent of X

Completeness of d-separation

By choosing the parametrization Θ carefully, we were able to establish an independence in the induced distribution which d-separation cannot detect.

Completeness of d-separation

By choosing the parametrization Θ carefully, we were able to establish an independence in the induced distribution which d-separation cannot detect.

If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$

then $\mathbf{X}$ and $\mathbf{Y}$ are independent given $\mathbf{Z}$ for any parametrization Θ

Completeness of d-separation

By choosing the parametrization Θ carefully, we were able to establish an independence in the induced distribution which d-separation cannot detect.

If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$

then $\mathbf{X}$ and $\mathbf{Y}$ are independent given $\mathbf{Z}$ for any parametrization Θ

If $\mathbf{X}$ and $\mathbf{Y}$ are not d-separated by $\mathbf{Z}$

then whether $\mathbf{X}$ and $\mathbf{Y}$ are dependent given $\mathbf{Z}$ depends on the specific parametrization Θ

Completeness of d-separation

Can we always parameterize a DAG G in such a way to ensure the completeness of d-separation?

Completeness of d-separation

Can we always parameterize a DAG G in such a way to ensure the completeness of d-separation?

For every DAG G, there is a parametrization Θ such that

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ if and only if $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

Completeness of d-separation

Can we always parameterize a DAG G in such a way to ensure the completeness of d-separation?

For every DAG G, there is a parametrization Θ such that

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ if and only if $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

There is no other graphical test

which can derive more independencies from $\mathrm{Markov}(G)$ than those derived by d-separation.

Further Properties of d-separation

Probabilistic independence does not satisfy Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Further Properties of d-separation

Probabilistic independence does not satisfy Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

d-separation satisfies Composition

$\mathrm{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $\mathrm{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $\mathrm{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Further Properties of d-separation

Probabilistic independence does not satisfy Composition

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

d-separation satisfies Composition

$\mathrm{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $\mathrm{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ only if $\mathrm{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Implication...

If we have a distribution that satisfies $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ but not $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$, there could not exist a DAG G which induces Pr and at the same time satisfies $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{W})$.

Further Properties of d-separation

d-separation satisfies additional properties

beyond Composition, which do not hold for arbitrary distributions.

Further Properties of d-separation

d-separation satisfies additional properties

beyond Composition, which do not hold for arbitrary distributions.

d-separation satisfies Intersection

$\text{dsep}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $\text{dsep}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $\text{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

Further Properties of d-separation

d-separation satisfies additional properties

beyond Composition, which do not hold for arbitrary distributions.

d-separation satisfies Intersection

$\text{dsep}(\mathbf{X}, \mathbf{Z} \cup \mathbf{W}, \mathbf{Y})$ and $\text{dsep}(\mathbf{X}, \mathbf{Z} \cup \mathbf{Y}, \mathbf{W})$ only if $\text{dsep}(\mathbf{X}, \mathbf{Z}, \mathbf{Y} \cup \mathbf{W})$

d-separation satisfies Chordality

$\text{dsep}(X, \{Z, W\}, Y)$ and $\text{dsep}(W, \{X, Y\}, Z)$ only if $\text{dsep}(X, Z, Y)$ or $\text{dsep}(X, W, Y)$

More on DAGs and Independence

G is an independence MAP (I-MAP) of $\Pr$ iff

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

More on DAGs and Independence

G is an independence MAP (I-MAP) of $\Pr$ iff

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

An I-MAP G is minimal iff

G ceases to be an I-MAP when we delete any edge from G

More on DAGs and Independence

G is an independence MAP (I-MAP) of $\Pr$ iff

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

An I-MAP G is minimal iff

G ceases to be an I-MAP when we delete any edge from G

By the semantics of Bayesian networks

if $\Pr$ is induced by a Bayesian network (G, Θ), then G must be an I-MAP of $\Pr$, although it may not be minimal.

More on DAGs and Independence

DAG G is a dependency MAP (D-MAP) of distribution Pr iff

$I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

More on DAGs and Independence

DAG G is a dependency MAP (D-MAP) of distribution $\Pr$ iff

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

If G is a D-MAP of $\Pr$

then the lack of d-separation in G implies a dependence in $\Pr$

More on DAGs and Independence

DAG G is a dependency MAP (D-MAP) of distribution $\Pr$ iff

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

If G is a D-MAP of $\Pr$

then the lack of d-separation in G implies a dependence in $\Pr$

DAG G is a perfect MAP (P-MAP) of distribution $\Pr$ iff

G is both an I-MAP and a D-MAP of $\Pr$

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)
- Consider the variables X_i one by one, for $i = 1, \ldots, n$

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)
- Consider the variables X_i one by one, for $i = 1, \ldots, n$
- For each variable X_i, identify a minimal subset $\mathbf{P}$ of the variables in $X_1, \ldots, X_{i-1}$ such that

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)
- Consider the variables X_i one by one, for $i = 1, \ldots, n$
- For each variable X_i, identify a minimal subset $\mathbf{P}$ of the variables in $X_1, \ldots, X_{i-1}$ such that
  - $I_{\Pr}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)
- Consider the variables X_i one by one, for $i = 1, \ldots, n$
- For each variable X_i, identify a minimal subset $\mathbf{P}$ of the variables in $X_1, \ldots, X_{i-1}$ such that
  - $I_{\Pr}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$
  - Make $\mathbf{P}$ the parents of X_i in DAG G

Minimal I-MAPs

Given a distribution $\Pr$, how can we construct a DAG G which is guaranteed to be a minimal I-MAP of $\Pr$?

Given an ordering $X_1, \ldots, X_n$ of the variables in $\Pr$:

- Start with an empty DAG G (no edges)
- Consider the variables X_i one by one, for $i = 1, \ldots, n$
- For each variable X_i, identify a minimal subset $\mathbf{P}$ of the variables in $X_1, \ldots, X_{i-1}$ such that
  - $I_{\Pr}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$
  - Make $\mathbf{P}$ the parents of X_i in DAG G

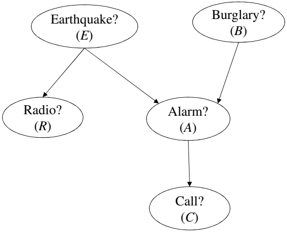
The resulting DAG is a minimal I-MAP of $\Pr$

Minimal I-MAPs

Construct a minimal I-MAP G for some distribution Pr using the previous procedure and variable order A, B, C, E, R.

Minimal I-MAPs

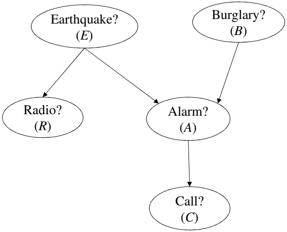
Construct a minimal I-MAP G for some distribution $\Pr$ using the previous procedure and variable order A, B, C, E, R.



Suppose that DAG G' is a P-MAP of distribution $\Pr$

Minimal I-MAPs

Construct a minimal I-MAP G for some distribution $\Pr$ using the previous procedure and variable order A, B, C, E, R.



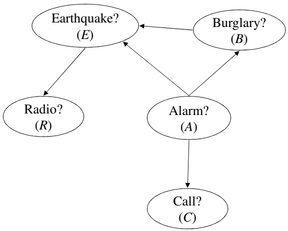
Suppose that DAG G' is a P-MAP of distribution $\Pr$

Independence tests $I_{\Pr}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$

can now be reduced to equivalent d-separation tests $\text{dsep}_{G'}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$

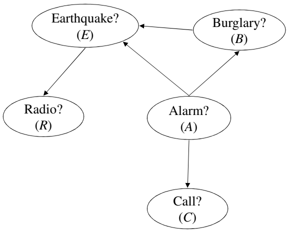
Minimal I-MAPs

Variable order A, B, C, E, R



Minimal I-MAPs

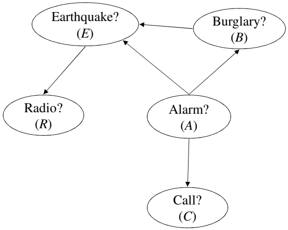
Variable order A, B, C, E, R



A: $\mathbf{P} = \emptyset$

Minimal I-MAPs

Variable order A, B, C, E, R

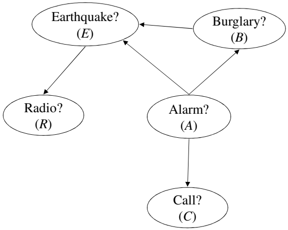


A: $\mathbf{P} = \emptyset$

B: $\mathbf{P} = A$ since $\text{dsep}_{G'}(B, A, \emptyset)$ and not $\text{dsep}_{G'}(B, \emptyset, A)$

Minimal I-MAPs

Variable order A, B, C, E, R



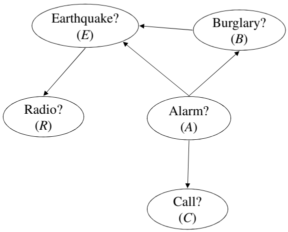
A: $\mathbf{P} = \emptyset$

B: $\mathbf{P} = A$ since $\text{dsep}_{G'}(B, A, \emptyset)$ and not $\text{dsep}_{G'}(B, \emptyset, A)$

C: $\mathbf{P} = A$ since $\text{dsep}_{G'}(C, A, B)$ and not $\text{dsep}(C, \emptyset, \{A, B\})$

Minimal I-MAPs

Variable order A, B, C, E, R



A: $\mathbf{P} = \emptyset$

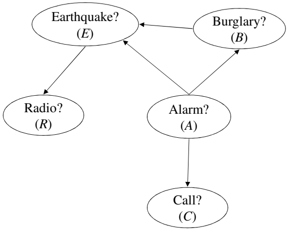
B: $\mathbf{P} = A$ since $\mathrm{dsep}_{G'}(B, A, \emptyset)$ and not $\mathrm{dsep}_{G'}(B, \emptyset, A)$

C: $\mathbf{P} = A$ since $\mathrm{dsep}_{G'}(C, A, B)$ and not $\mathrm{dsep}(C, \emptyset, \{A, B\})$

E: $\mathbf{P} = A, B$ is the smallest subset of A, B, C such that $\mathrm{dsep}_{G'}(E, \mathbf{P}, \{A, B, C\} \setminus \mathbf{P})$

Minimal I-MAPs

Variable order A, B, C, E, R



A: $\mathbf{P} = \emptyset$

B: $\mathbf{P} = A$ since $\text{dsep}_{G'}(B, A, \emptyset)$ and not $\text{dsep}_{G'}(B, \emptyset, A)$

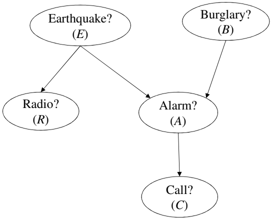
C: $\mathbf{P} = A$ since $\text{dsep}_{G'}(C, A, B)$ and not $\text{dsep}(C, \emptyset, \{A, B\})$

E: $\mathbf{P} = A, B$ is the smallest subset of A, B, C such that $\text{dsep}_{G'}(E, \mathbf{P}, \{A, B, C\} \setminus \mathbf{P})$

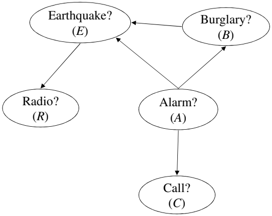
R: $\mathbf{P} = E$ is the smallest subset of A, B, C, E such that $\text{dsep}_{G'}(R, \mathbf{P}, \{A, B, C, E\} \setminus \mathbf{P})$

Minimal I-MAPs

DAG G' and distribution $\Pr$

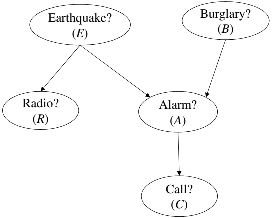


Minimal I-MAP G

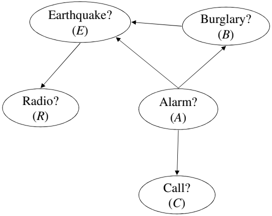


Minimal I-MAPs

DAG G' and distribution Pr



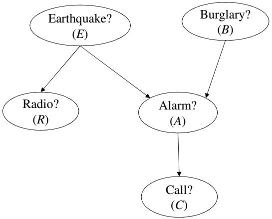
Minimal I-MAP G



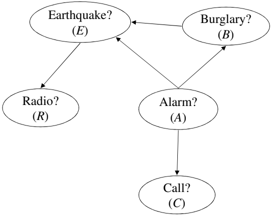
- If $\mathrm{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, then $\mathrm{dsep}_{G'}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$

Minimal I-MAPs

DAG G' and distribution $\Pr$



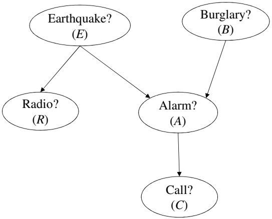
Minimal I-MAP G



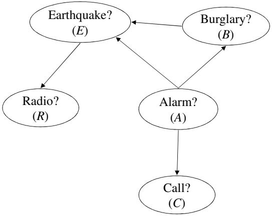
- If $\text{dsep}_G(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, then $\text{dsep}_{G'}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ and $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$
- This ceases to hold if we delete any of the five edges in G

Minimal I-MAPs

DAG G' and distribution Pr



Minimal I-MAP G



- If $\mathrm{dsep}_G(\mathbf{X},\mathbf{Z},\mathbf{Y})$, then $\mathrm{dsep}_{G'}(\mathbf{X},\mathbf{Z},\mathbf{Y})$ and $I_{\mathrm{Pr}}(\mathbf{X},\mathbf{Z},\mathbf{Y})$
- This ceases to hold if we delete any of the five edges in G

If we delete the edge $E \leftarrow B$

we will have $\mathrm{dsep}_G(E, A, B)$, yet $\mathrm{dsep}_{G'}(E, A, B)$ does not hold.

Uniqueness of Minimal I-MAPs

The minimal I-MAP of a distribution is not unique

we may get different ones depending on the chosen variable order.

Uniqueness of Minimal I-MAPs

The minimal I-MAP of a distribution is not unique

we may get different ones depending on the chosen variable order.

Even when using the same variable ordering

we may have multiple minimal subsets $\mathbf{P}$ of $\{X_1, \ldots, X_{i-1}\}$ for which $I_{\mathrm{Pr}}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$

Uniqueness of Minimal I-MAPs

The minimal I-MAP of a distribution is not unique

we may get different ones depending on the chosen variable order.

Even when using the same variable ordering

we may have multiple minimal subsets $\mathbf{P}$ of $\{X_1, \ldots, X_{i-1}\}$ for which $I_{\mathrm{Pr}}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$

This can only happen if

the probability distribution represents some logical constraints.

Uniqueness of Minimal I-MAPs

The minimal I-MAP of a distribution is not unique

we may get different ones depending on the chosen variable order.

Even when using the same variable ordering

we may have multiple minimal subsets $\mathbf{P}$ of $\{X_1, \ldots, X_{i-1}\}$ for which $I_{\mathrm{Pr}}(X_i, \mathbf{P}, \{X_1, \ldots, X_{i-1}\} \setminus \mathbf{P})$

This can only happen if

the probability distribution represents some logical constraints.

We can ensure the uniqueness of a minimal I-MAP for a given variable ordering

if we restrict ourselves to strictly positive distributions.

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

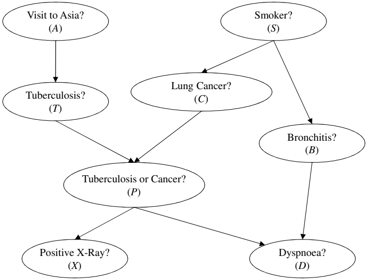
If distribution $\Pr$ is induced by DAG G

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

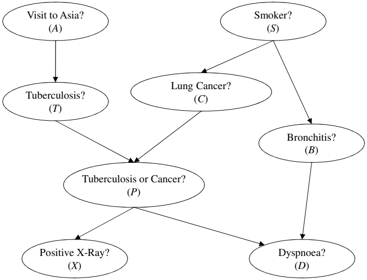
the two variables have a common child in DAG G

Blankets and Boundaries



Markov blanket for C

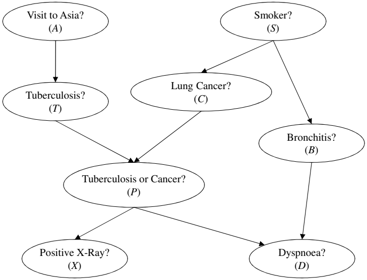
Blankets and Boundaries

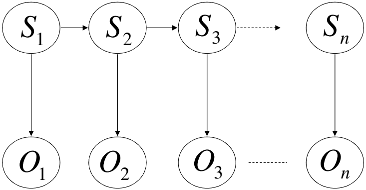


Markov blanket for C

S, P, T

Blankets and Boundaries



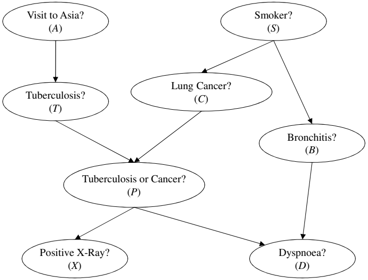


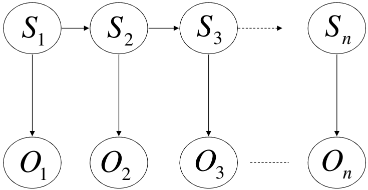
Markov blanket for S_t, $t > 1$

Markov blanket for C

S, P, T

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