



Machine learning: lecture 20

Tommi S. Jaakkola

MIT CSAIL

tommi@csail.mit.edu



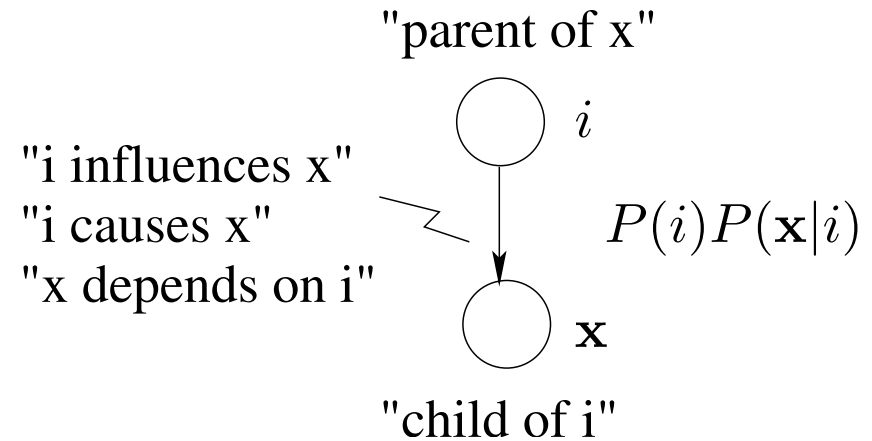
Outline

- Bayesian networks
 - examples, specification
 - graphs and independence
 - associated distribution

Bayesian networks

- Bayesian networks are directed acyclic graphs, where the nodes represent variables and directed edges capture dependencies

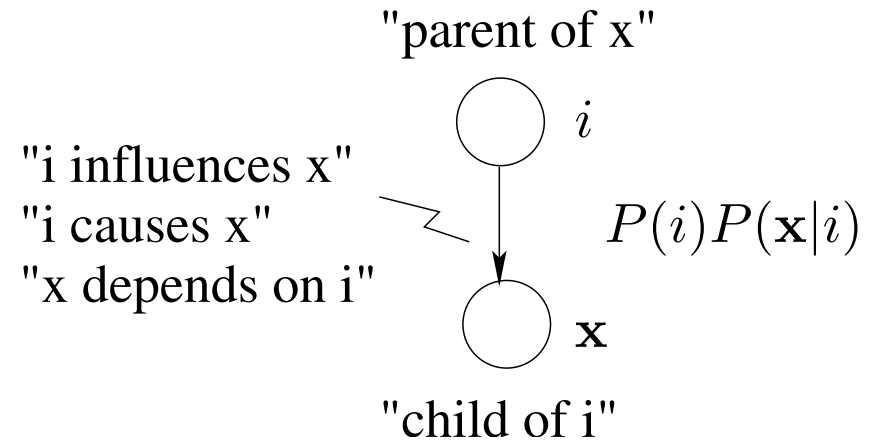
A mixture model as a Bayesian network



Bayesian networks

- Bayesian networks are directed acyclic graphs, where the nodes represent variables and directed edges capture dependencies

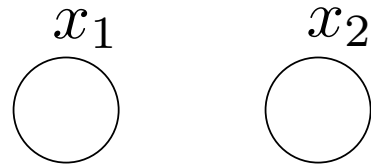
A mixture model as a Bayesian network



- Graph semantics:
 - graph \Rightarrow separation properties \Rightarrow independence
- Association with probability distributions:
 - independence \Rightarrow family of distributions

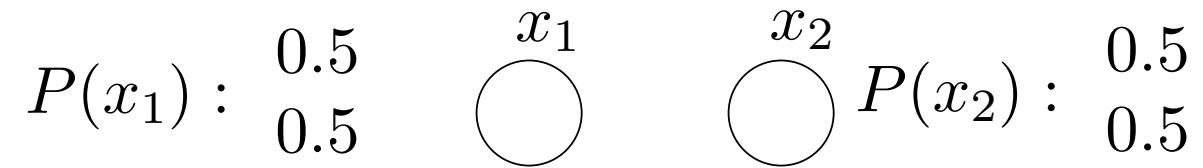
Example

- A simple Bayesian network: coin tosses



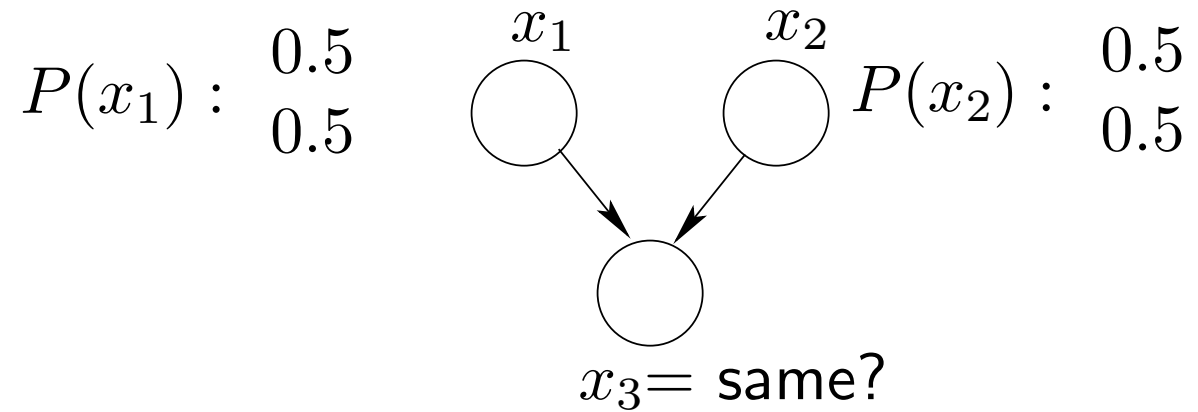
Example

- A simple Bayesian network: coin tosses



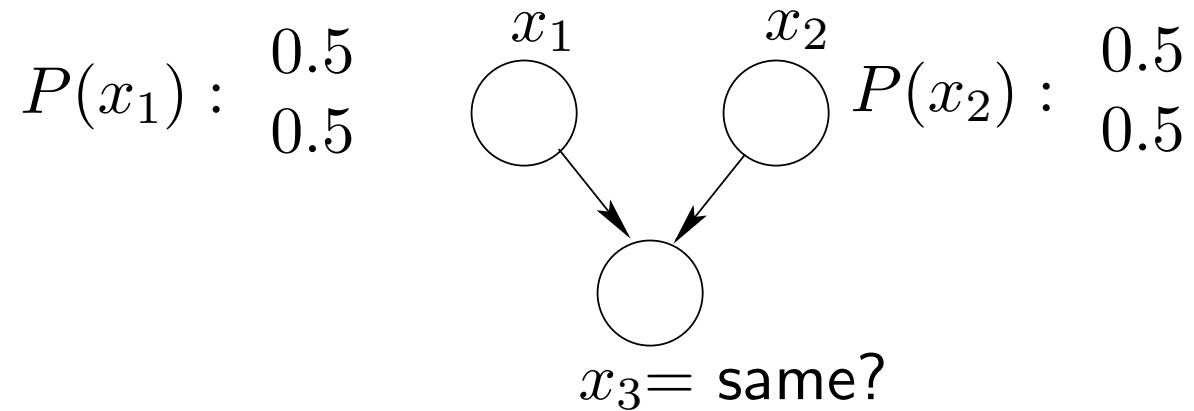
Example

- A simple Bayesian network: coin tosses



Example

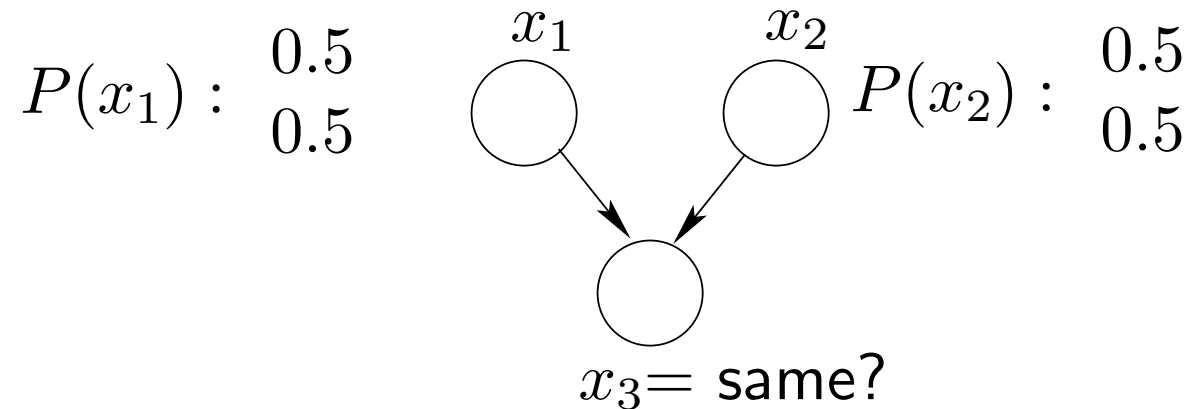
- A simple Bayesian network: coin tosses



		<i>hh</i>	<i>ht</i>	<i>th</i>	<i>tt</i>
$P(x_3 x_1, x_2) :$	<i>y</i>	1.0	0.0	0.0	1.0
	<i>n</i>	0.0	1.0	1.0	0.0

Example

- A simple Bayesian network: coin tosses

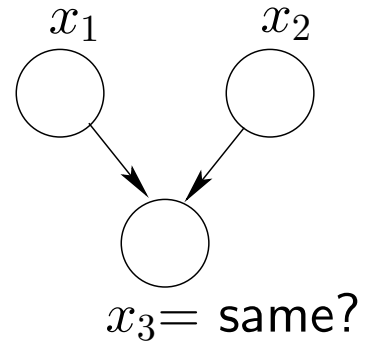


		<i>hh</i>	<i>ht</i>	<i>th</i>	<i>tt</i>
$P(x_3 x_1, x_2) :$	<i>y</i>	1.0	0.0	0.0	1.0
	<i>n</i>	0.0	1.0	1.0	0.0

- Two levels of description
 1. graph structure (dependencies, independencies)
 2. associated probability distribution

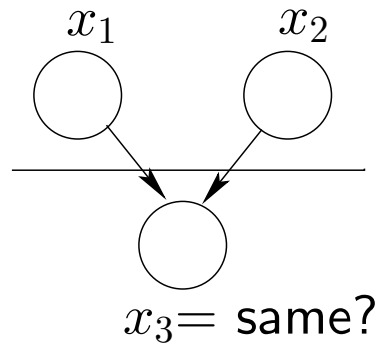
Example cont'd

- What can the graph alone tell us?



Example cont'd

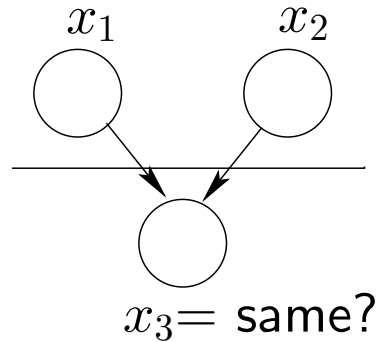
- What can the graph alone tell us?



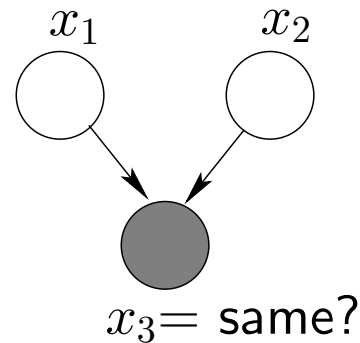
- x_1 and x_2 are *marginally independent*

Example cont'd

- What can the graph alone tell us?



- x_1 and x_2 are *marginally independent*



- x_1 and x_2 become *dependent* if we know x_3
(the dependence concerns our beliefs about the outcomes)

Traffic example

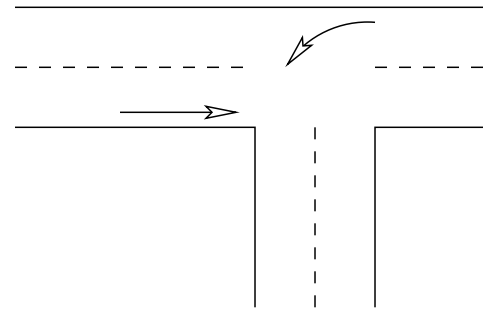
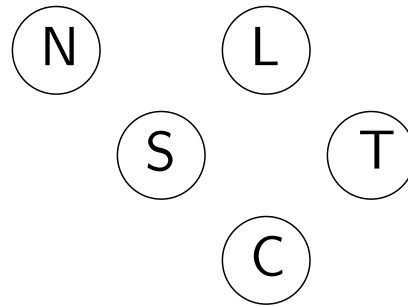
N = X is nice?

L = traffic light

S = X decides to stop?

T = the other car turns left?

C = crash?



Traffic example

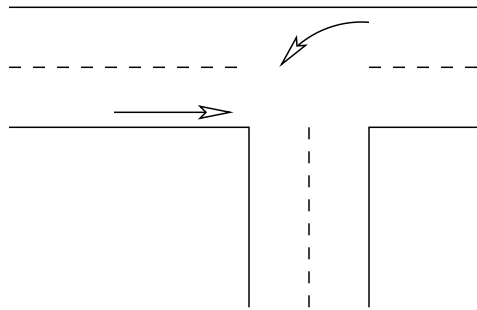
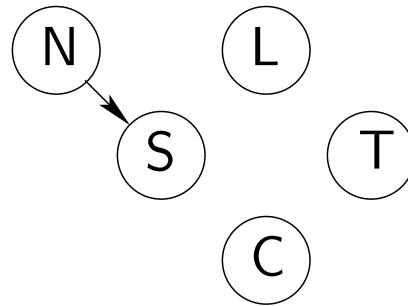
N = X is nice?

L = traffic light

S = X decides to stop?

T = the other car turns left?

C = crash?



Traffic example

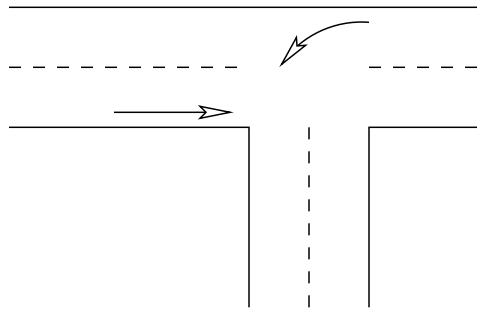
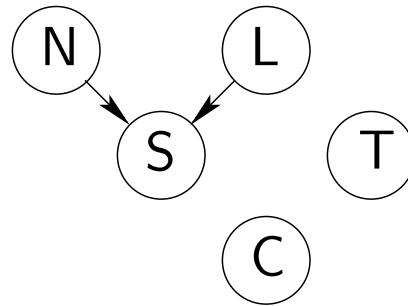
N = X is nice?

L = traffic light

S = X decides to stop?

T = the other car turns left?

C = crash?



Traffic example

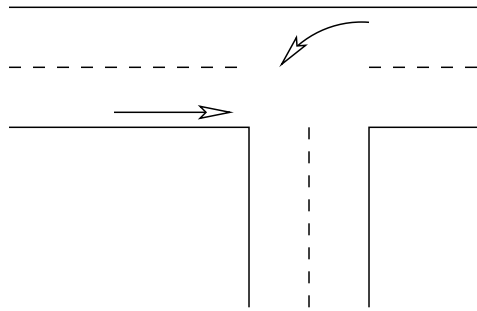
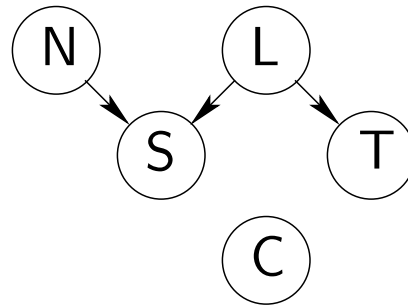
N = X is nice?

L = traffic light

S = X decides to stop?

T = the other car turns left?

C = crash?



Traffic example

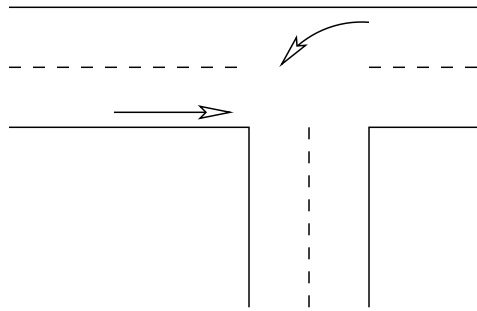
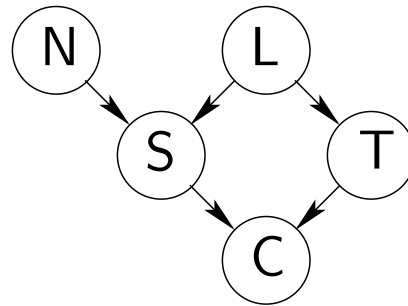
N = X is nice?

L = traffic light

S = X decides to stop?

T = the other car turns left?

C = crash?



Traffic example

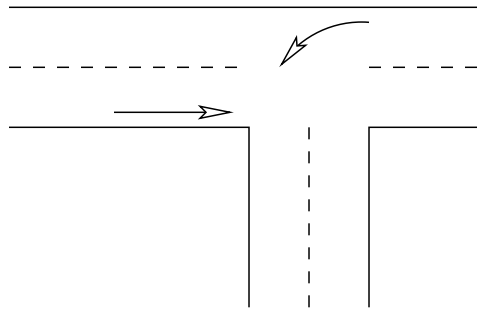
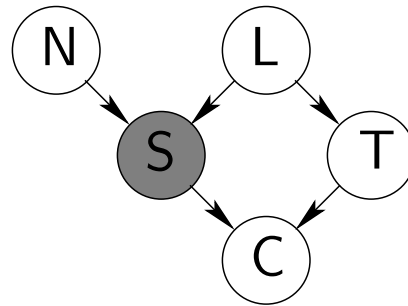
N = X is nice?

L = traffic light

S = X decides to stop?

T = the other car turns left?

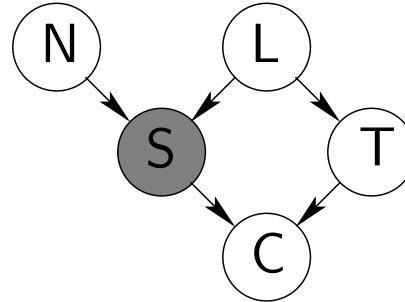
C = crash?



- If we only know that X decided to stop, can X's character (variable N) tell us anything about the other car turning (variable T)?

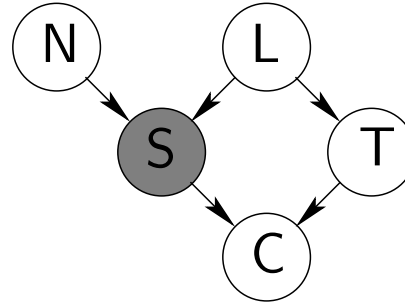
Graph, independence, d-separation

- Are N and T independent given S ?



Graph, independence, d-separation

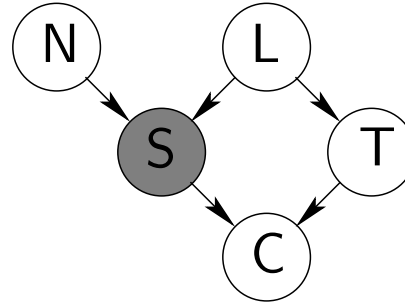
- Are N and T independent given S ?



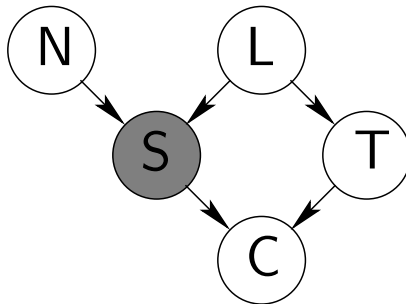
Definition: Variables N and T are D-separated given S if S separates them in the moralized ancestral graph

Graph, independence, d-separation

- Are N and T independent given S ?



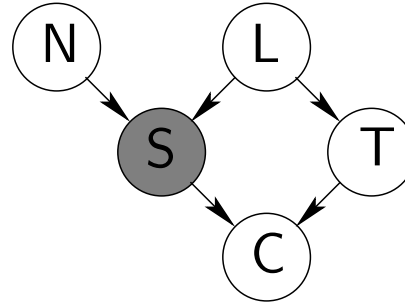
Definition: Variables N and T are D-separated given S if S separates them in the moralized ancestral graph



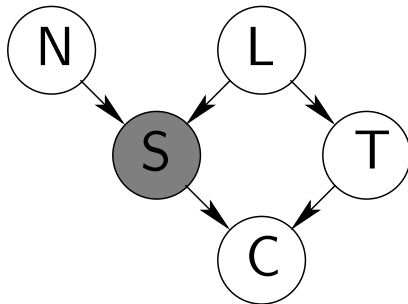
original

Graph, independence, d-separation

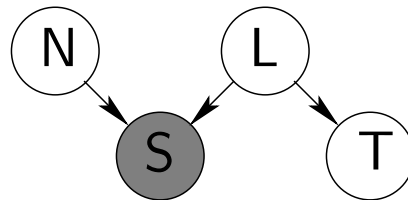
- Are N and T independent given S ?



Definition: Variables N and T are D-separated given S if S separates them in the moralized ancestral graph



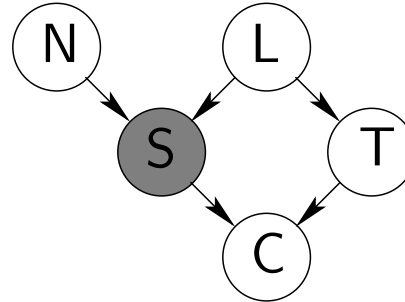
original



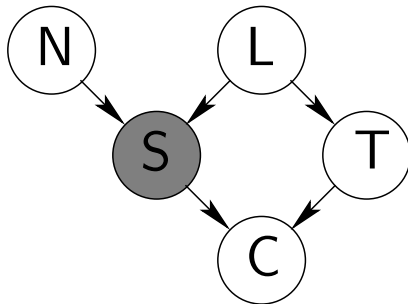
ancestral

Graph, independence, d-separation

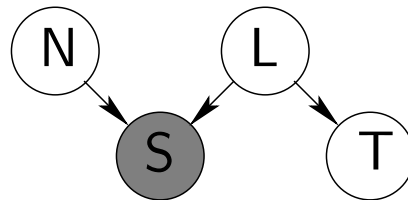
- Are N and T independent given S ?



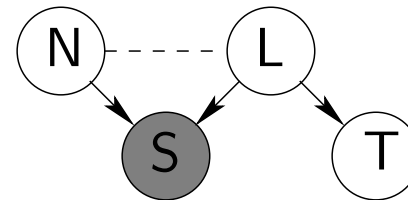
Definition: Variables N and T are D-separated given S if S separates them in the moralized ancestral graph



original



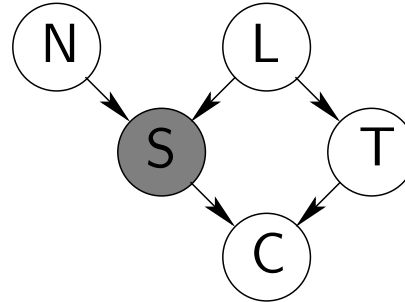
ancestral



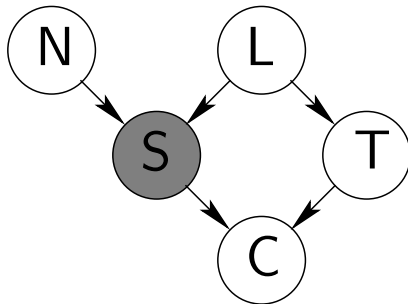
moralized ancestral

Graph, independence, d-separation

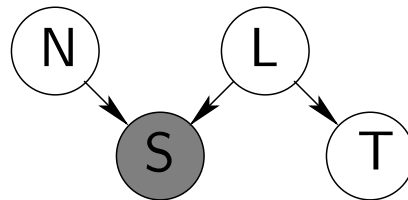
- Are N and T independent given S ?



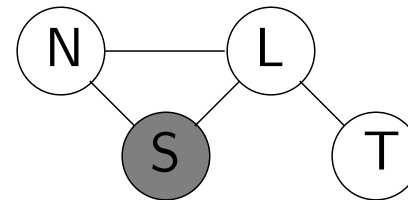
Definition: Variables N and T are D-separated given S if S separates them in the moralized ancestral graph



original

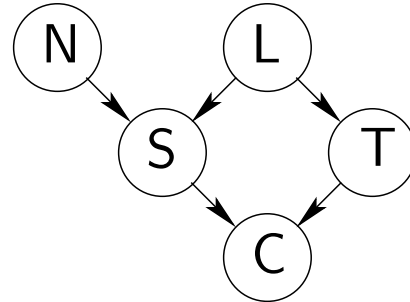


ancestral



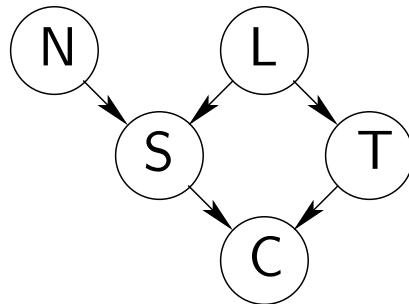
moralized ancestral

Graphs and distributions



- A graph is a compact representation of a large collection of independence properties

Graphs and distributions



- A graph is a compact representation of a large collection of independence properties

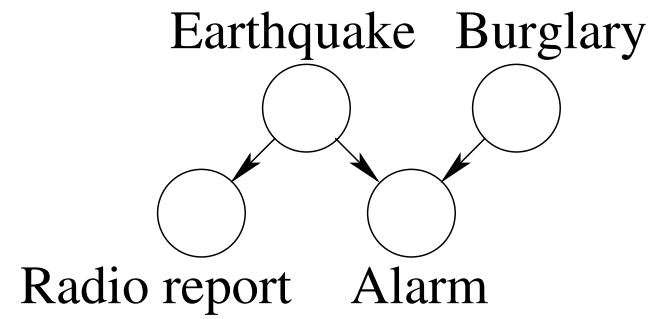
Theorem: Any probability distribution that is consistent with a directed graph G has to factor according to “node given parents”:

$$P(\mathbf{x}|G) = \prod_{i=1}^d P(x_i|\mathbf{x}_{pa_i})$$

where \mathbf{x}_{pa_i} are the *parents* of x_i and d is the number of nodes (variables) in the graph.

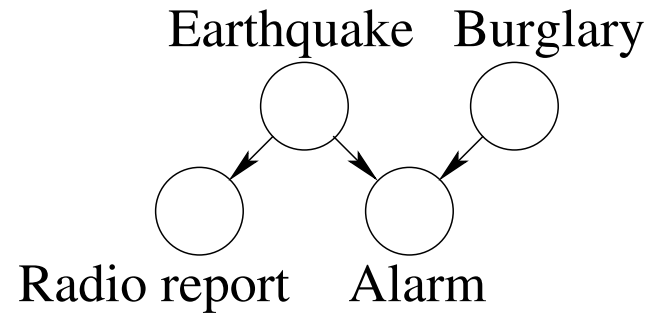
Explaining away phenomenon

- Model

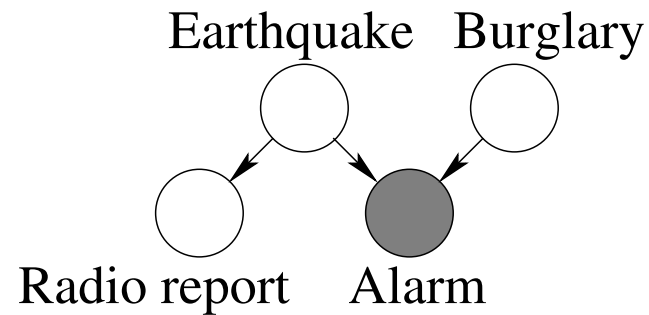


Explaining away phenomenon

- Model

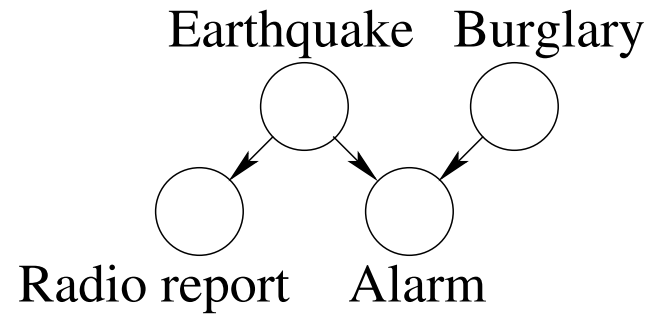


- Evidence, competing causes

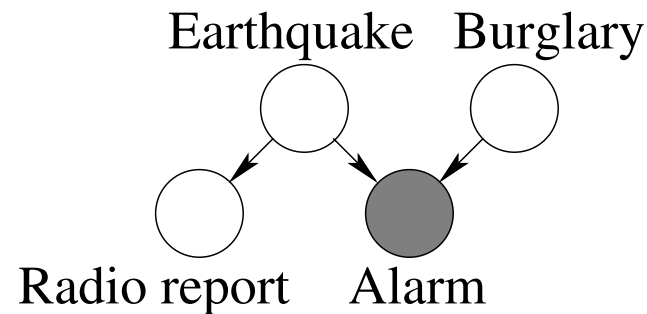


Explaining away phenomenon

- Model



- Evidence, competing causes



- Additional evidence and explaining away

