

Bayesian Network

Julyedu: Johnson

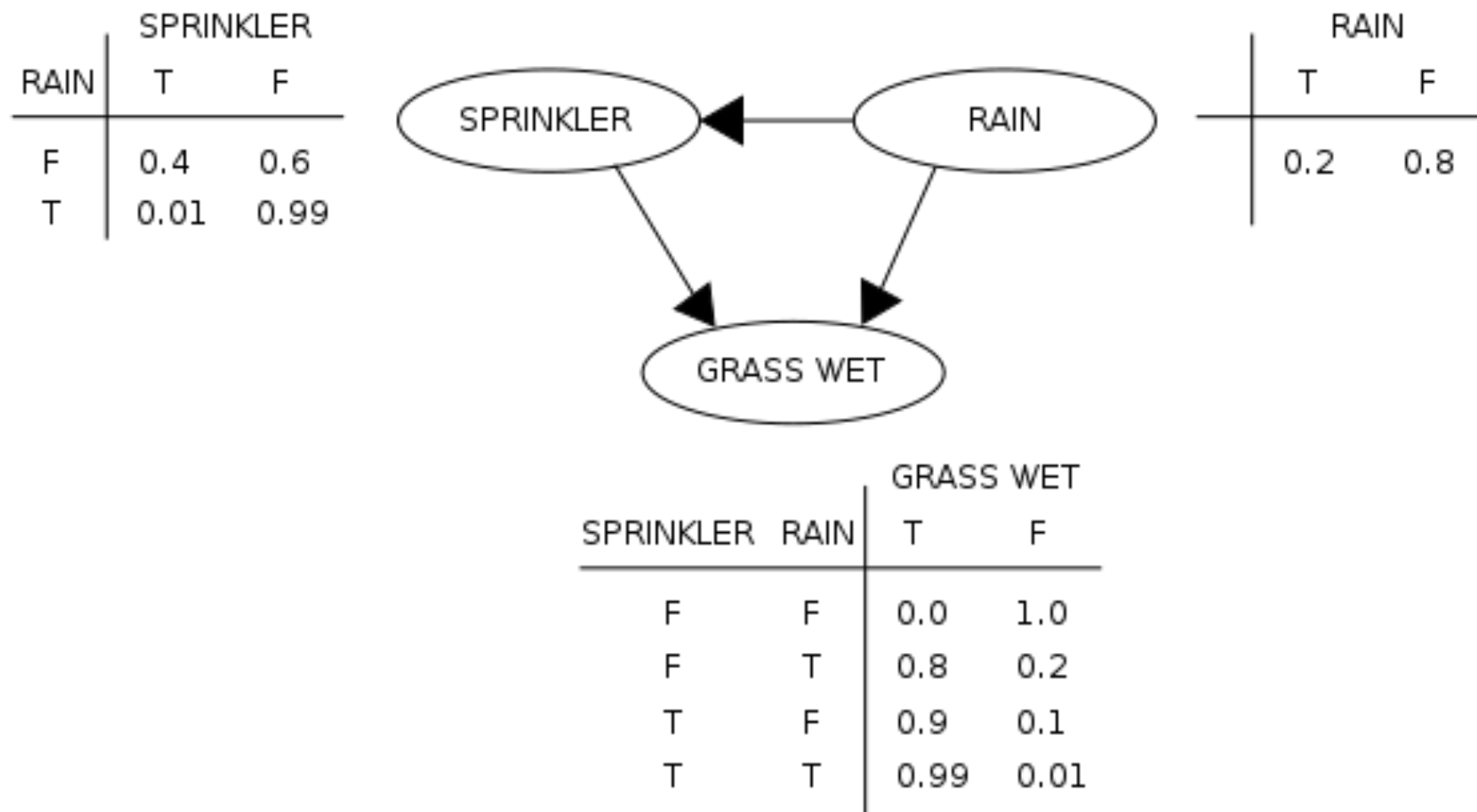
2018/07/31

Bayesian Network

- What is Bayesian Network?
- Examples:
 - Naive Bayes (朴素贝叶斯)
 - Hidden Markov Model (隐马尔可夫模型)
 - Latent Dirichlet Allocation (LDA, 主题模型)
- Main Topic:
 - Probability theory + Graph theory
 - Probabilistic graphical model

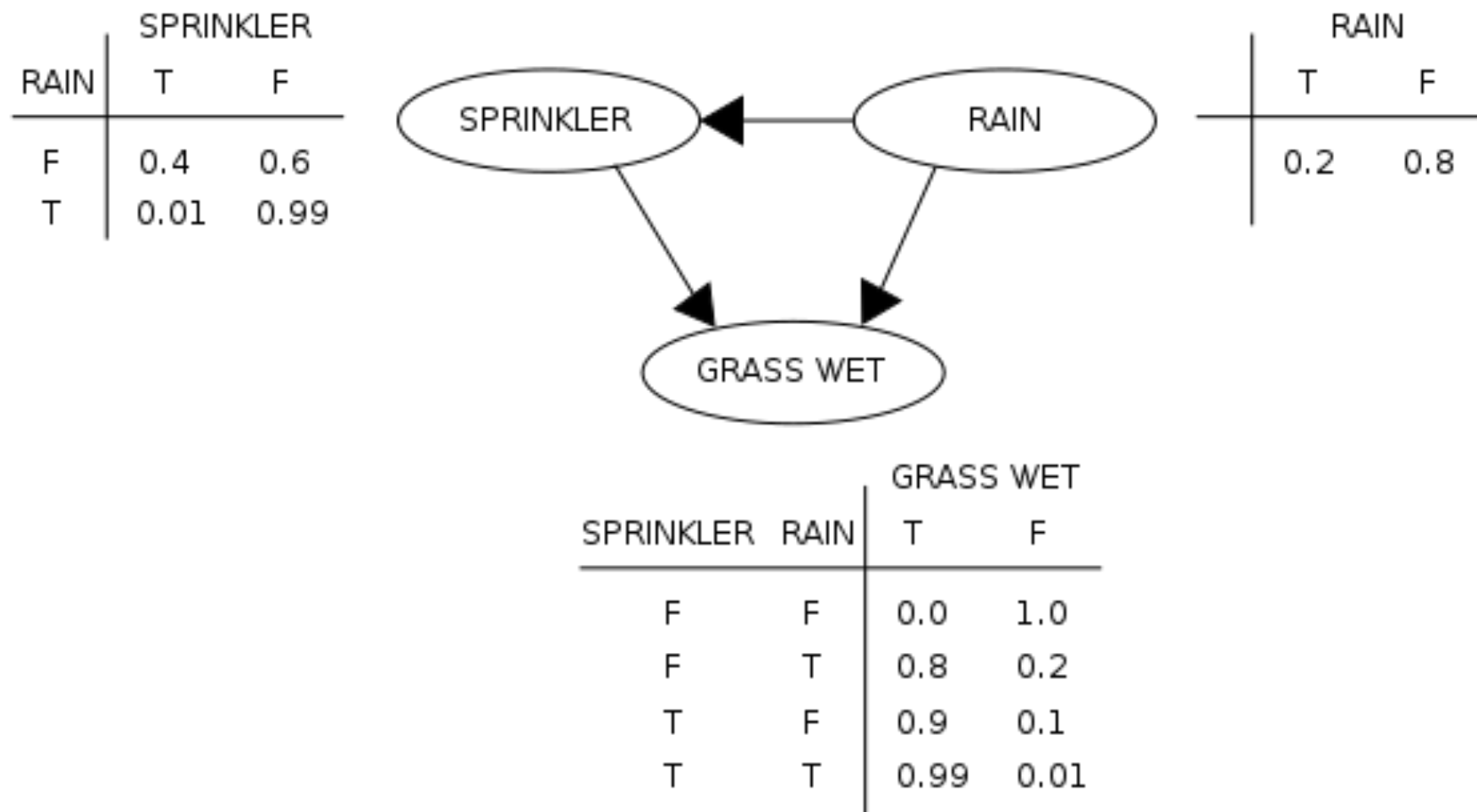
Examples

- Use directed graph to represent conditional independency:



Examples

- Use directed graph to represent conditional independency:



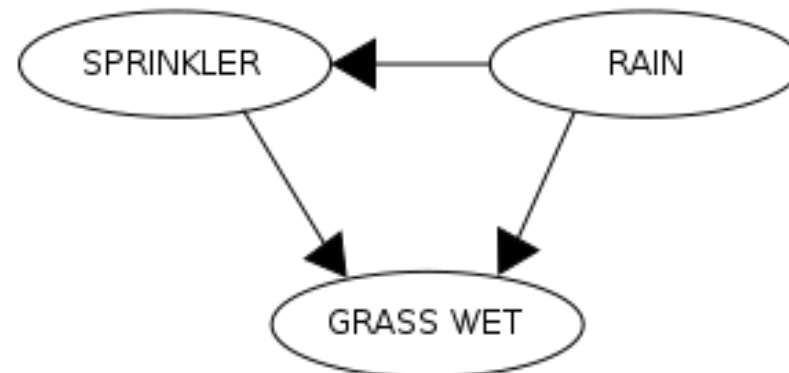
- Each node corresponds to a random variable.
- Each edge corresponds to a conditional dependency between variables.

Definition

- A **Bayesian network** is specified by a directed acyclic graph:
 - $G = (V, E)$ where:
 - One node i for each random variable V_i
 - One conditional probability distribution (CPD) per node:

$$\Pr(V_i \mid V_{pa(i)})$$

RAIN	SPRINKLER	
	T	F
F	0.4	0.6
T	0.01	0.99



	RAIN	
	T	F
	0.2	0.8

SPRINKLER	RAIN	GRASS WET	
		T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Definition

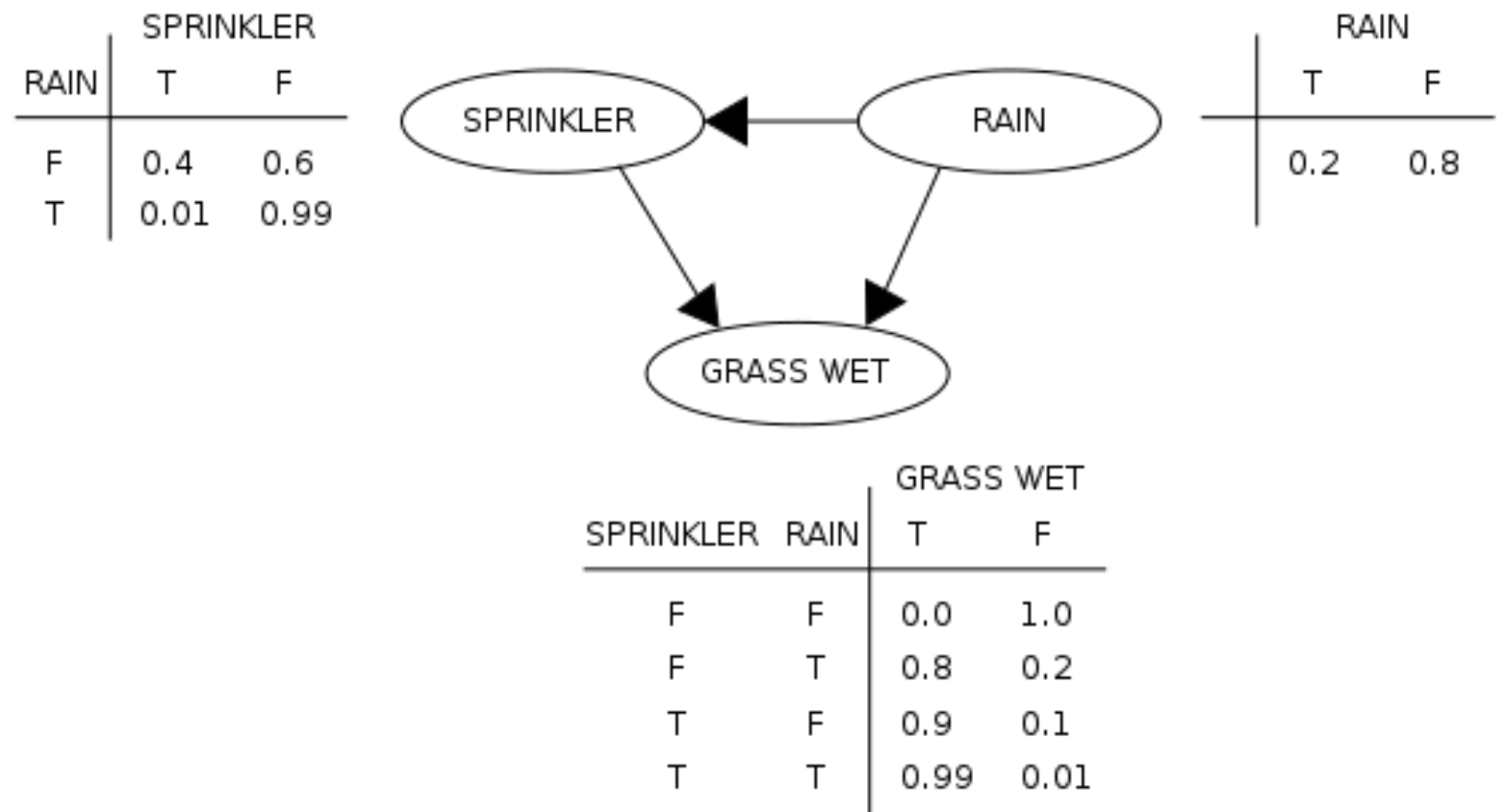
- A **Bayesian network** is specified by a directed acyclic graph:
 - $G = (V, E)$ where:
 - One node i for each random variable V_i
 - One conditional probability distribution (CPD) per node:
 $\Pr(V_i \mid V_{pa(i)})$
- Graph structure specifies the factorization of the joint distribution:

$$\Pr(V_1, \dots, V_n) = \prod_{i \in [n]} \Pr(V_i \mid V_{pa(i)})$$

Definition

- Graph structure specifies the factorization of the joint distribution:

$$\Pr(V_1, \dots, V_n) = \prod_{i \in [n]} \Pr(V_i \mid V_{pa(i)})$$



Definition

- Graph structure specifies the factorization of the joint distribution:

$$\Pr(V_1, \dots, V_n) = \prod_{i \in [n]} \Pr(V_i \mid V_{pa(i)})$$

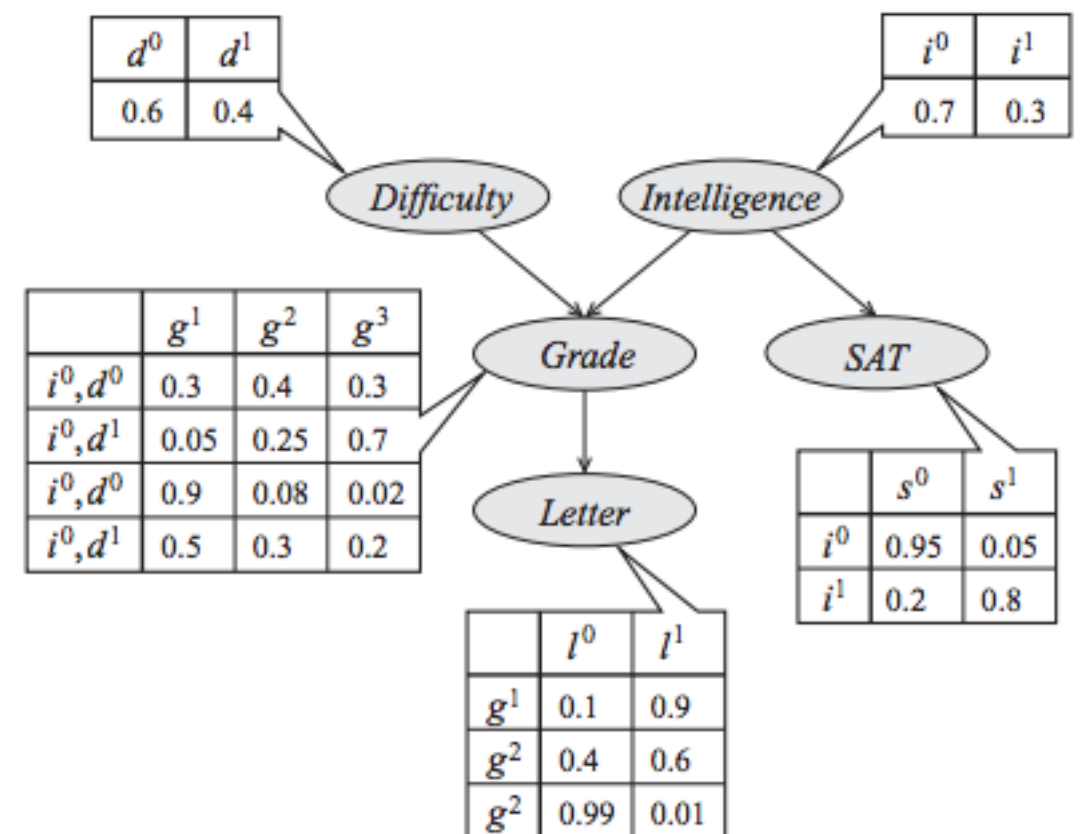
- What's the advantages of Bayesian networks?
 - Compact representation of a probability distribution
 - Clear probabilistic semantics
 - Direct graph can be used to represent causal relationship

More Examples

- Graph structure specifies the factorization of the joint distribution:

$$\Pr(V_1, \dots, V_n) = \prod_{i \in [n]} \Pr(V_i \mid V_{pa(i)})$$

- What's the joint distribution?

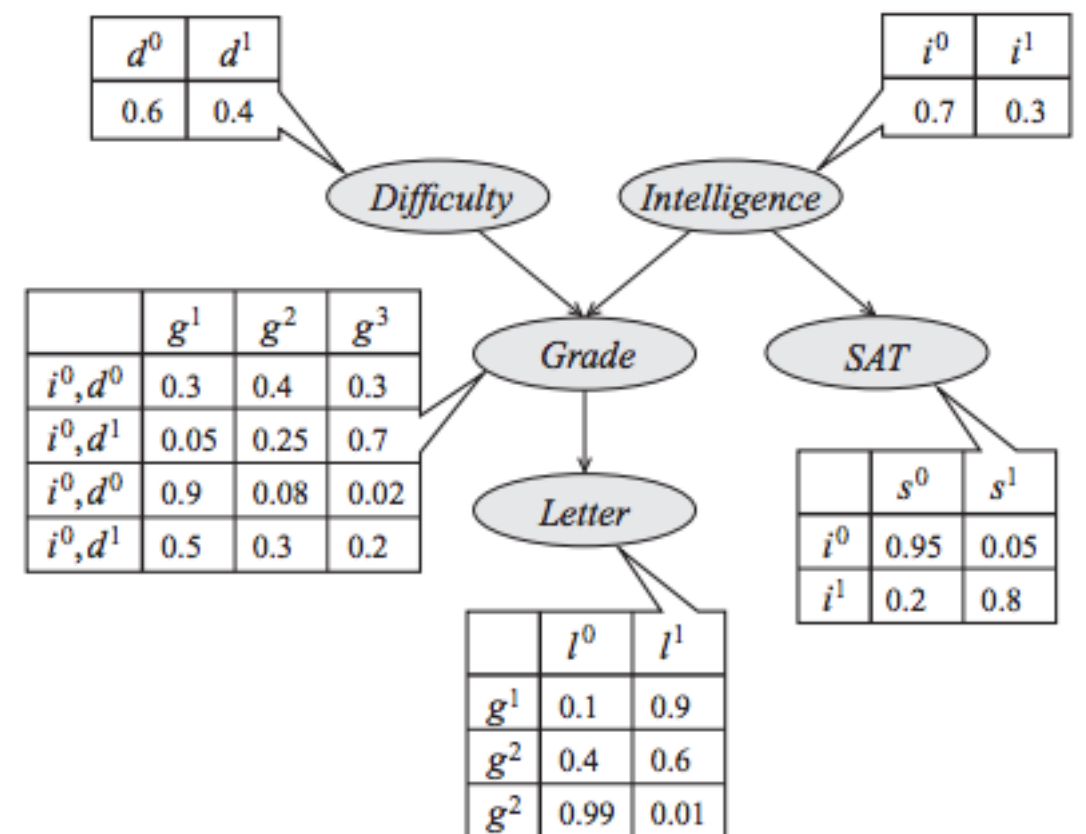


More Examples

- Graph structure specifies the factorization of the joint distribution:

$$\Pr(V_1, \dots, V_n) = \prod_{i \in [n]} \Pr(V_i \mid V_{pa(i)})$$

- What's the joint distribution?



$$\Pr(d, i, g, s, l) = \Pr(d) \cdot \Pr(i) \cdot \Pr(g \mid d, i) \cdot \Pr(s \mid i) \cdot \Pr(l \mid g)$$

Conditional Independence

- Graph structure implies conditional independency
- If two variables are conditionally independent, graph has no edge between them:

$$D \perp I$$

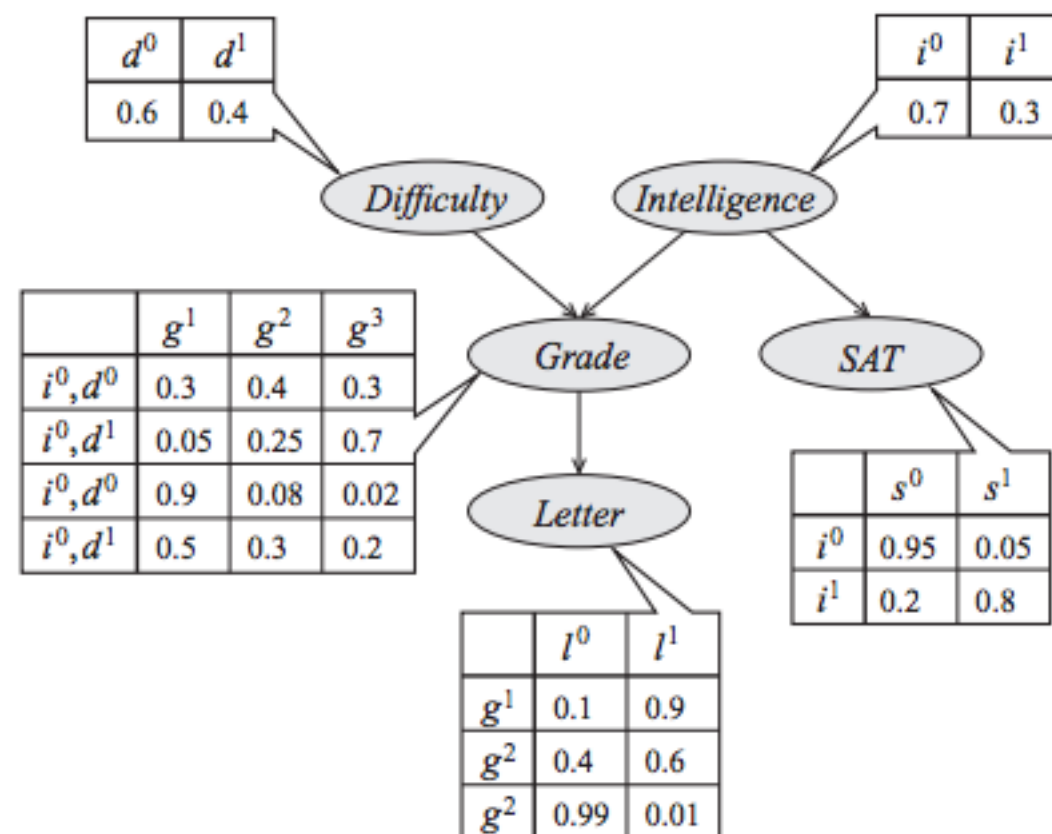
$$L \perp S \mid I$$

$$G \perp S \mid I$$

$$D \perp S$$

$$D \perp L \mid G$$

$$L \perp S \mid G$$




Probabilistic Graphical Models



acm


MORE ACM AWARDS



A.M. TURING AWARD

A.M. TURING CENTENARY CELEBRATION WEBCAST





A.M. TURING AWARD WINNERS BY...

ALPHABETICAL LISTING

YEAR OF THE AWARD

RESEARCH SUBJECT



JUDEA PEARL 

United States – 2011

CITATION


For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.



SHORT ANNOTATED BIBLIOGRAPHY



ACM TURING AWARD LECTURE VIDEO



RESEARCH SUBJECTS



ADDITIONAL MATERIALS



Photo-Essay

BIRTH:

September 4, 1936, Tel Aviv.

EDUCATION:

D.S., Electrical Engineering (Technion, 1960); M.S., Electronics (Newark College of Engineering, 1961); M.S., Physics (Rutgers University, 1965); Ph.D., Electrical Engineering (Polytechnic Institute of Brooklyn, 1965).

EXPERIENCE:

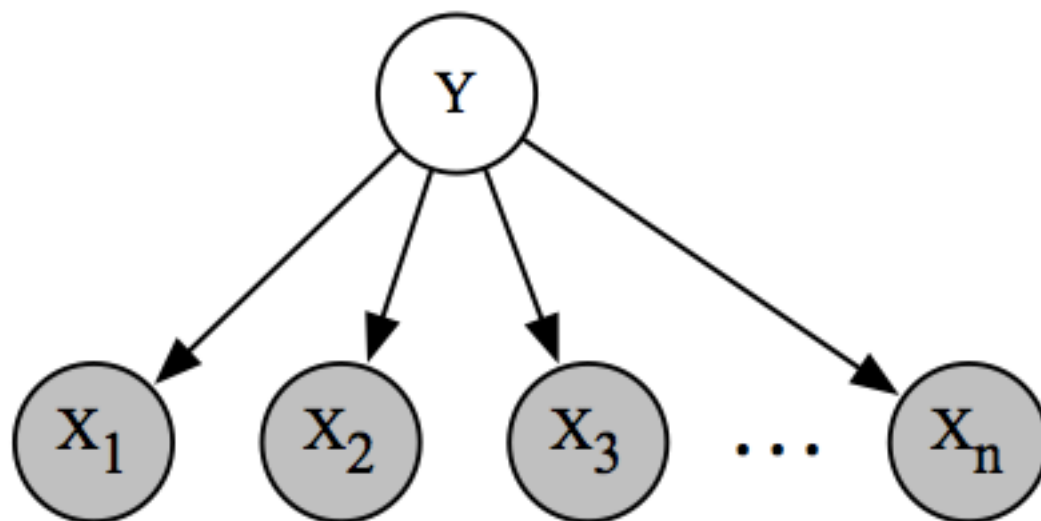
Judea Pearl created the representational and computational foundation for the processing of information under uncertainty.

He is credited with the invention of *Bayesian networks*, a mathematical formalism for defining complex probability models, as well as the principal algorithms used for inference in these models. This work not only revolutionized the field of artificial intelligence but also became an important tool for many other branches of engineering and the natural sciences. He later created a mathematical framework for *causal inference* that has had significant impact in the social sciences.

Judea Pearl was born on September 4, 1936, in Tel Aviv, which was at that time administered under the British Mandate for Palestine. He grew up in *Bnei Brak*, a Biblical town his grandfather went to reestablish in 1924. In 1956, after serving in the Israeli army and joining a Kibbutz, Judea decided to study engineering. He attended

More Examples

- Naive Bayes as a Bayesian network:

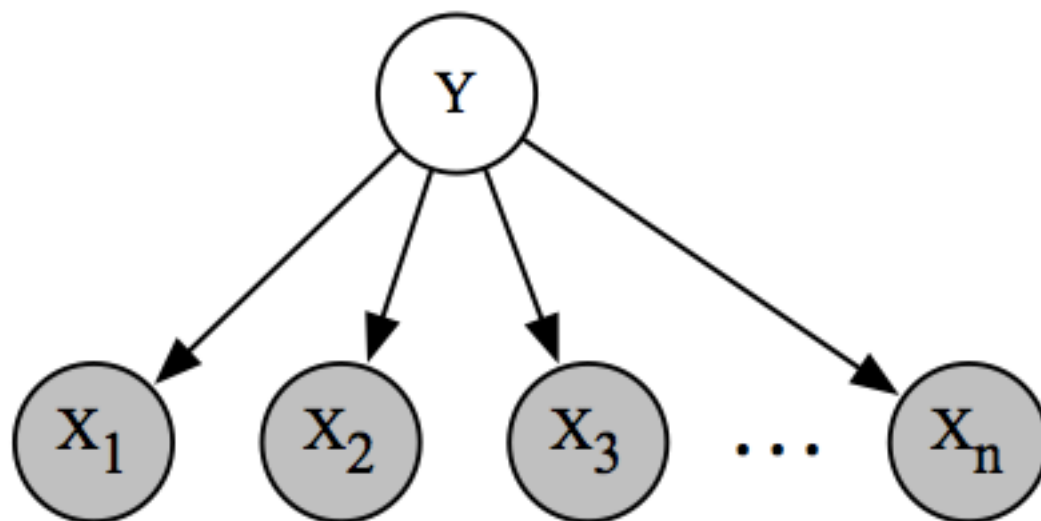


$$\Pr(Y, X_1, \dots, X_n) = \Pr(Y) \prod_{i \in [n]} \Pr(X_i \mid Y)$$

- Shaded nodes are **observed**
- White nodes are **hidden**

More Examples

- Naive Bayes as a Bayesian network:



$$\Pr(Y, X_1, \dots, X_n) = \Pr(Y) \prod_{i \in [n]} \Pr(X_i \mid Y)$$

- Problem: given **observed**, infer **hidden** ?

More Examples

- Problem: given **observed**, infer **hidden** ?

GAUSSIAN
NAIVE BAYES
CLASSIFIER

"Gaussian" because this is a normal distribution →

This is our prior belief →

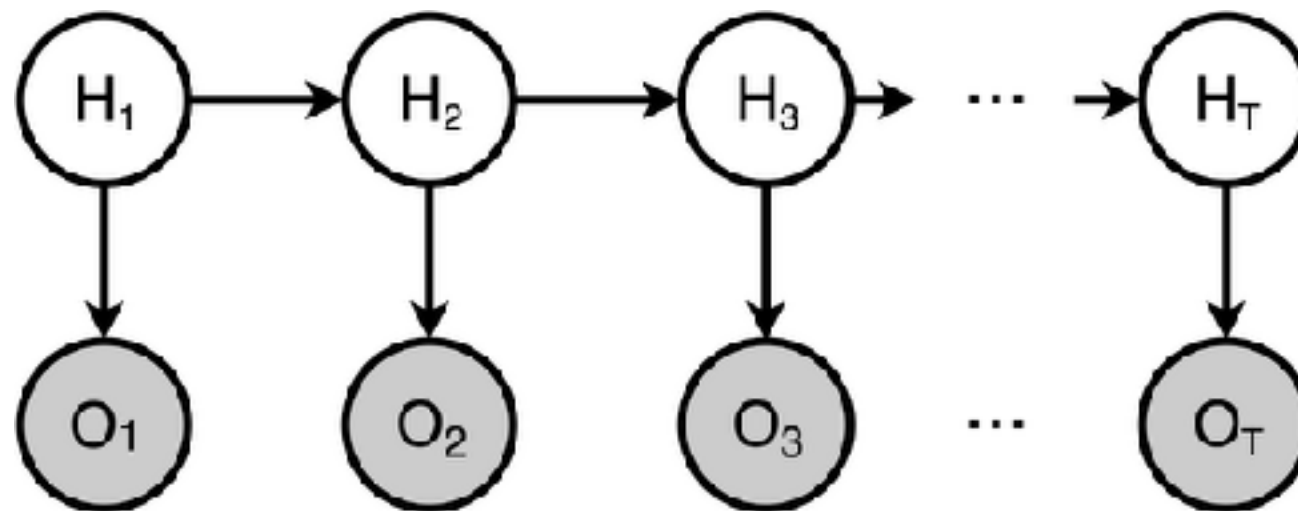
$$P(\text{class} | \text{data}) = \frac{P(\text{data} | \text{class}) \times P(\text{class})}{P(\text{data})}$$

We don't calculate this in naive bayes classifiers →

ChrisAlbon

More Examples

- Hidden Markov model as a Bayesian network:

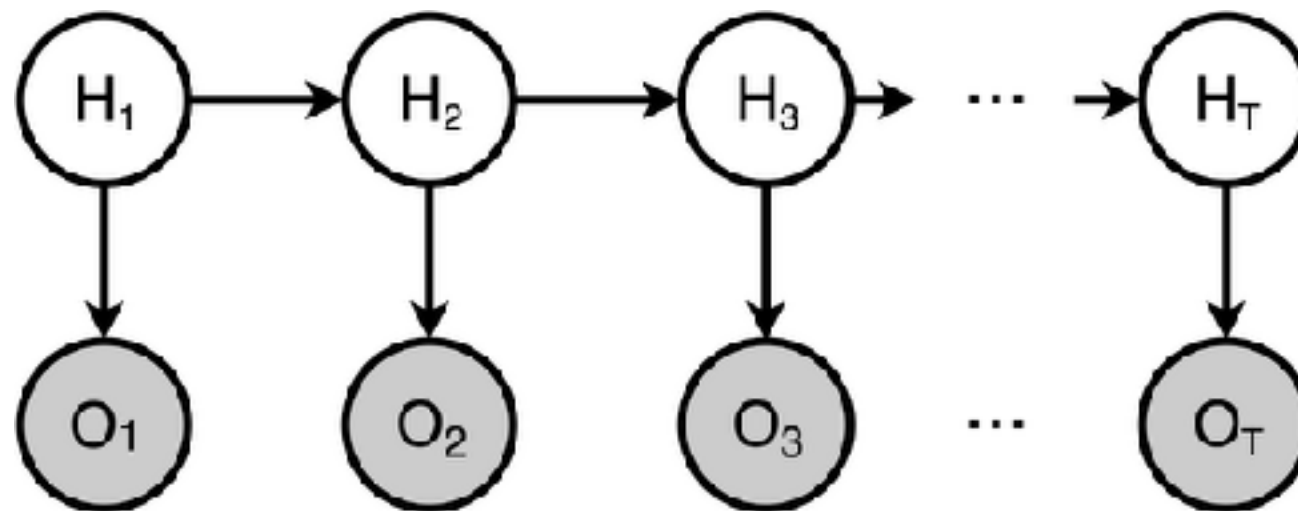


$$\Pr(H_1, \dots, H_T, O_1, \dots, O_T) = \Pr(H_1) \prod_{i \in [T]} \Pr(O_i \mid H_i) \prod_{i \in [T-1]} \Pr(H_{i+1} \mid H_i)$$

- Shadowed nodes are **observed**
- White nodes are **hidden**

More Examples

- Hidden Markov model as a Bayesian network:

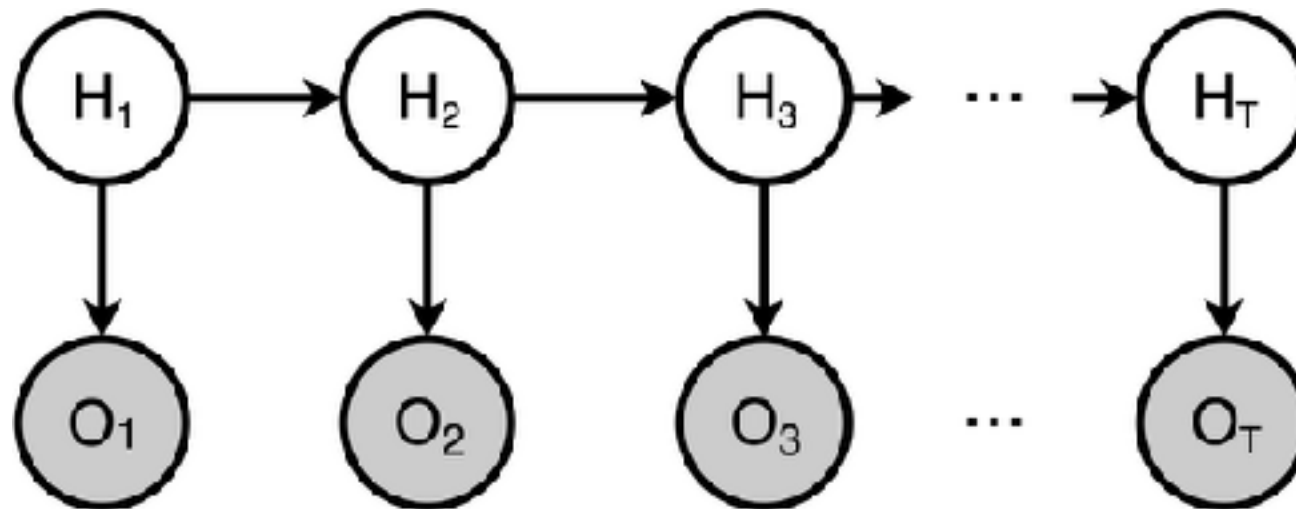


$$\Pr(H_1, \dots, H_T, O_1, \dots, O_T) = \Pr(H_1) \prod_{i \in [T]} \Pr(O_i \mid H_i) \prod_{i \in [T-1]} \Pr(H_{i+1} \mid H_i)$$

- Applications:
 - POS Tagging (词性标注)
 - Speech recognition (语音识别)
 - Word segmentation (中文分词)
 -

More Examples

- Hidden Markov model as a Bayesian network:



$$\Pr(H_1, \dots, H_T, O_1, \dots, O_T) = \Pr(H_1) \prod_{i \in [T]} \Pr(O_i \mid H_i) \prod_{i \in [T-1]} \Pr(H_{i+1} \mid H_i)$$

- Algorithms:
 - Forward-Backward algorithm (inference)
 - Expectation maximization algorithm (learning)

Conditional Independence

- Review:
 - Independence:

$$\Pr(A, B) = \Pr(A) \cdot \Pr(B)$$

$$\Pr(A \mid B) = \Pr(A)$$

$$\Pr(B \mid A) = \Pr(B)$$

How to show the above three formulas are equivalent?

Conditional Independence

- Review:
 - Conditional independence:

$$\Pr(A, B \mid C) = \Pr(A \mid C) \cdot \Pr(B \mid C)$$

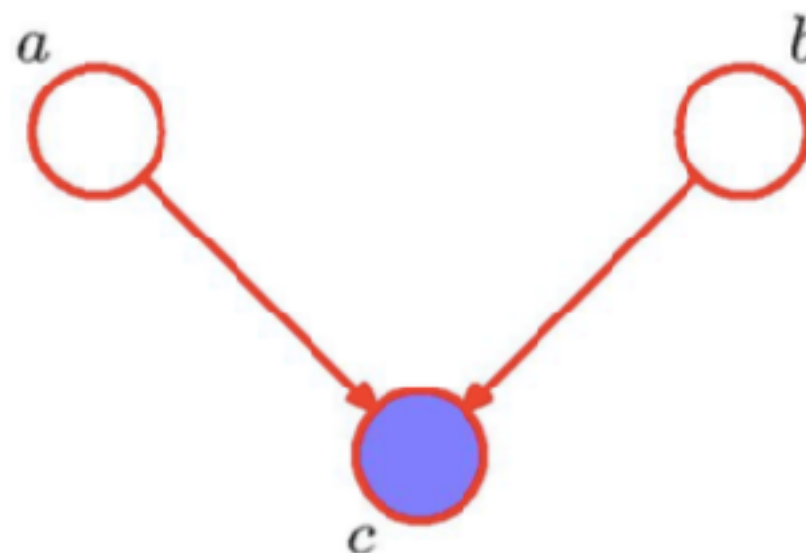
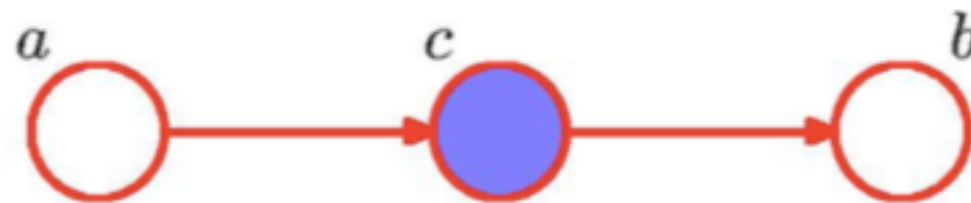
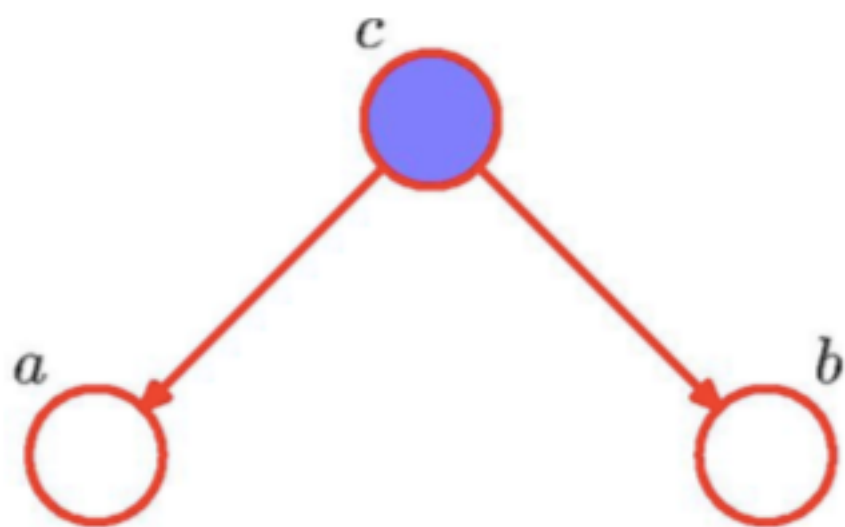
$$\Pr(A \mid B, C) = \Pr(A \mid C)$$

$$\Pr(B \mid A, C) = \Pr(B \mid C)$$

How to show the above three formulas are equivalent?

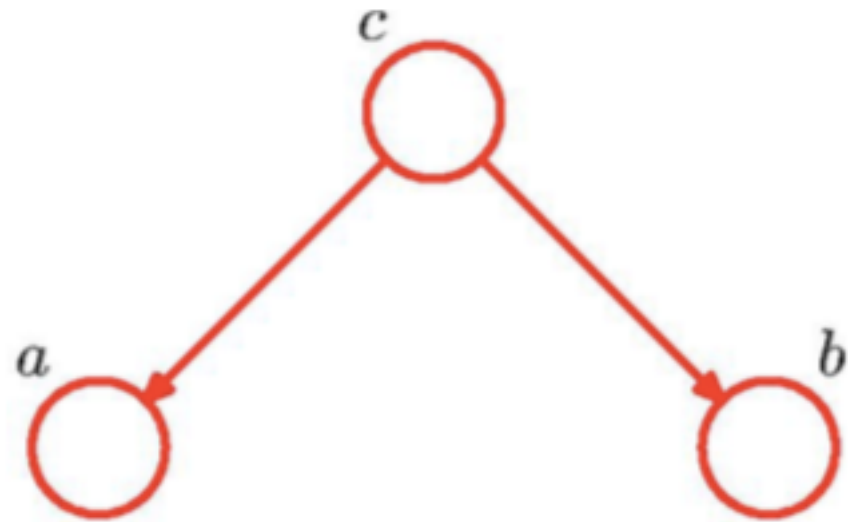
Conditional Independence

- **d-separation:**
 - A method to quickly judge conditional independence given the graph
 - 3 difference cases

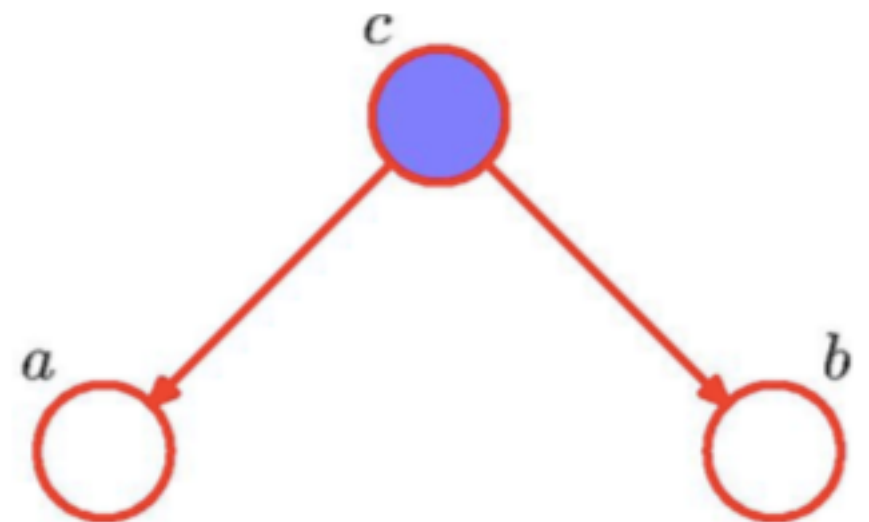


Conditional Independence

- Tail-to-Tail:



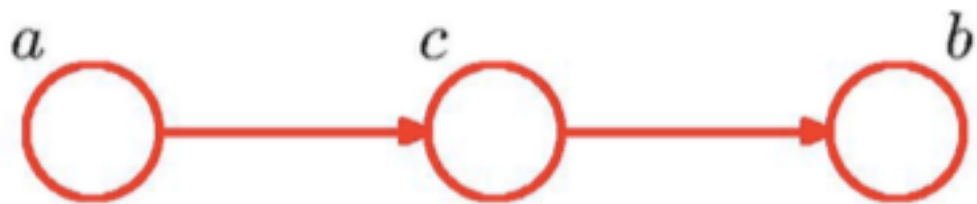
$$a \not\perp b$$



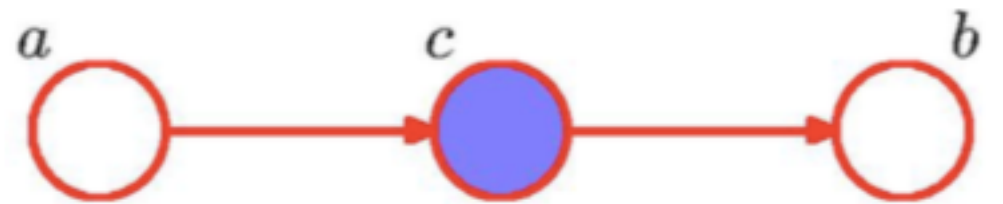
$$a \perp b \mid c$$

Conditional Independence

- **Head-to-Tail:**



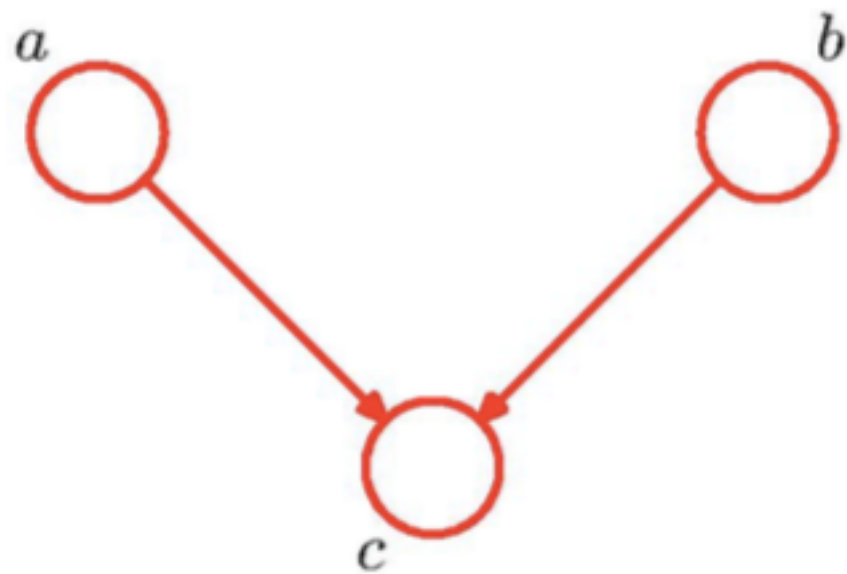
$$a \not\perp b$$



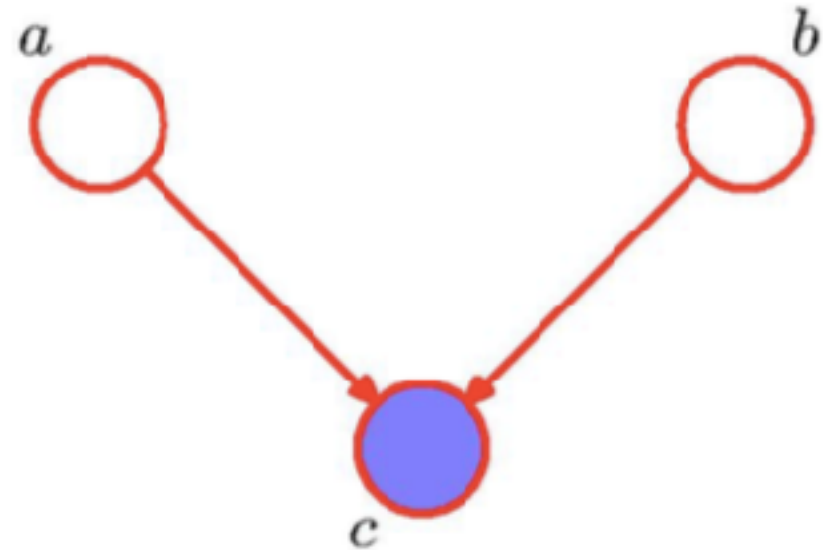
$$a \perp b \mid c$$

Conditional Independence

- Head-to-Head:



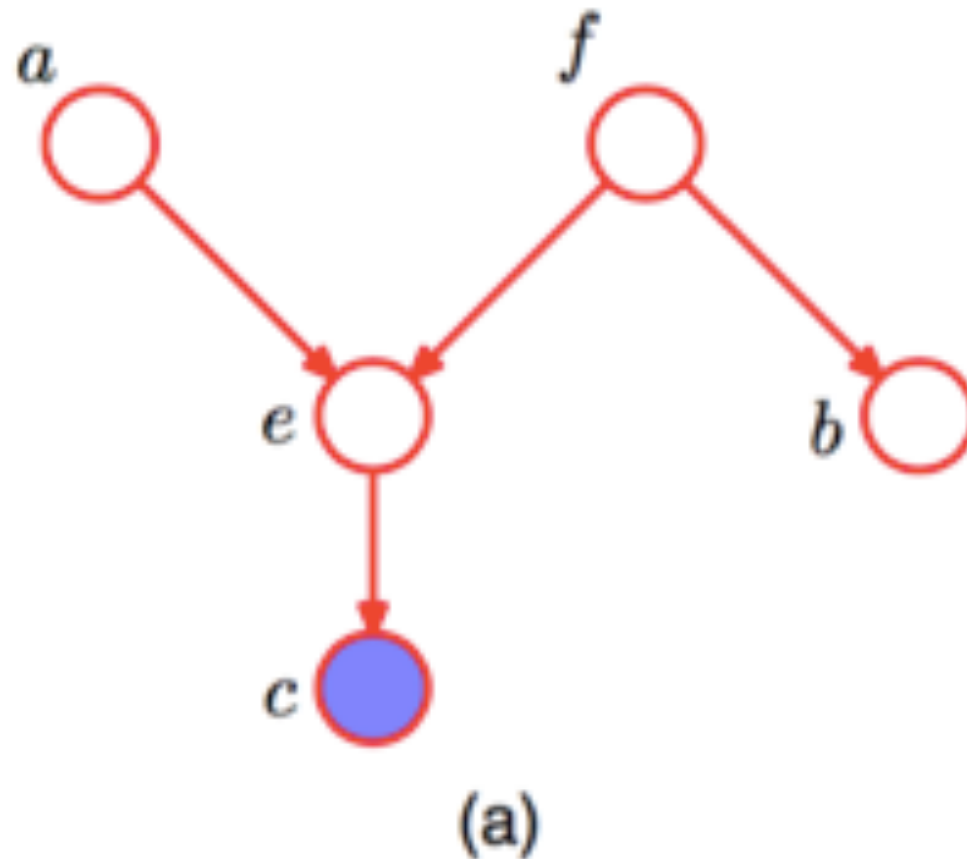
$$a \perp b$$



$$a \perp b \mid c$$

Conditional Independence

- Examples:



- Is:

$$a \perp b \mid c?$$

$$a \perp f \mid c?$$

$$e \perp b?$$

Thanks and Questions