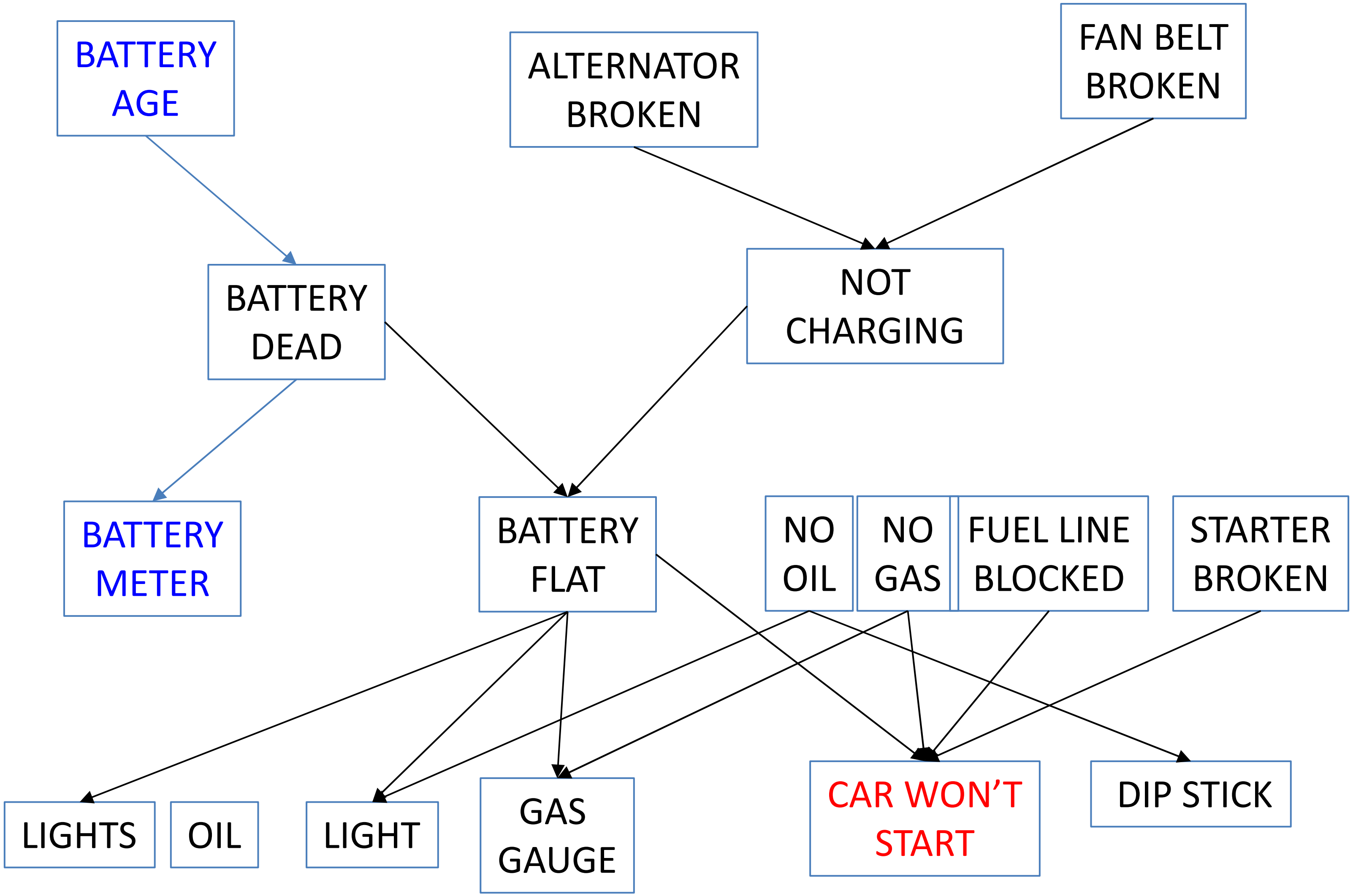
Artificial Intelligence Fundamentals

Learning: Bayes Network

Example – Bayes Network



Can we diagnose the problem?

- Bayes Network
  - Composed of nodes
  - Nodes correspond to events
  - You might or might not know the events (random variables)
  - Nodes are linked by arcs
  - Arcs suggest that a child node is influenced (in a probabilistic way) by its parent
  - 16 variables
  - The space is 2^{16} possible values
- The Bayes Network is a compact representation of a distribution over this very large joint probability distribution of all of these variables
  - Once we specify it, we can observe (e.g., the car won't start and the lights, etc.) and compute probability (like the alternator is broken)

Bayes Network

Used extensively in almost all fields of smart computer systems

Google

diagnostics

finance

prediction

robotics

machine learning

Building blocks of more advanced AI techniques

Particle Filters

Kalman Filters

Hidden Markov Models

many others

MDP’s and POMDP’s

Probabilities

- Used to express uncertainty
- Coin
  - Probability for heads is 0.5 $P(H) = \frac{1}{2}$
  - Probability for tails is $P(T) = \frac{1}{2}$
  - $P(T) + P(H) = 1$
  - Suppose $P(T) = \frac{1}{4}$, $P(H) = \frac{3}{4}$
  - Suppose $P(H) = \frac{1}{2}$, the probability having three heads into a row will be: $P(H, H, H) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$, the coin flips are independent

Probabilities

- Flip the coin 4 times
  - X_i=result of i-*th* coin flip $X_i = \{H, T\}, P_i(H) = \frac{1}{2}$
  - $P(X_1 = X_2 = X_3 = X_4) =?$
  - 1/16 + 1/16 = 1/8
  - $P(X_1 X_2 X_3 X_4\ contains\ at\ least\ 3H) =?$
  - 1/16+1/16+1/16+1/16+1/16=5/16

Probabilities

- Basic
  - $0 \leq P(A) \leq 1$
  - $P(True) = 1$ and $P(False\ ) = 0$
- Complementary probability (for binary event)
  - $P(A) = p \Rightarrow P(\neg A) = 1 - p$
- Independence, X independent of Y ($X \perp Y$)
  - $X \perp Y : P(X)P(Y) = P(X, Y)$
    - P(X),P(Y) – marginals
    - P(X,Y) – joint probability
  - $X \perp Y : P(X|Y) = P(X)$

Dependence

- $P(X_1 = H) = \frac{1}{2}, \begin{cases} H: P(X_2 = H | X_1 = H) = 0.9 \\ T: P(X_2 = T | X_1 = T) = 0.8 \end{cases}$
- $P(X_2 = H)$=?
- $P(X_2 = H) = P(X_2 = H | X_1 = H) \cdot P(X_1 = H) + P(X_2 = H | X_1 = T) \cdot P(X_1 = T) = 0.9 \cdot 0.5 + 0.2 \cdot 0.5 = 0.45 + 0.1 = 0.55$

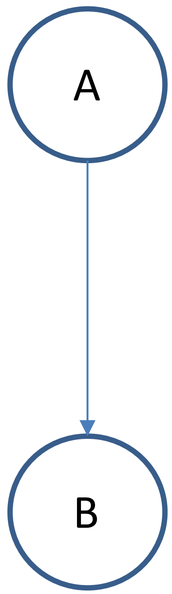
- Total probability:
  – $P(Y) = \sum_i P(Y | X = i) \cdot P(X = i)$
- Negation of probabilities
  – $P(\neg X | Y) = 1 - P(X | Y)$
- Joint probability
  – $P(X, Y) = P(X | Y) \cdot P(Y)$
- Chain rule
  – $P(X_n, X_{n-1}, \ldots, X_1) = \prod_{i=n}^{1} P(X_i | X_{i-1}, \ldots, X_1) = P(X_n | X_{n-1}, X_{n-2}, \ldots, X_1) P(X_{n-1} | X_{n-2}, \ldots, X_1) \ldots P(X_2 | X_1) P(X_1)$

Bayes Rule

- $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$
- Cancer example:
  - $P(C) = 0.01$, $P(\neg C) = 0.99$
  - $P(+|C) = 0.9$, $P(-|C) = 0.1$
  - $P(+|\neg C) = 0.2$, $P(-|\neg C) = 0.8$
  - $P(C|+) = \frac{P(+|C) \cdot P(C)}{P(+)} = \frac{P(+|C) \cdot P(C)}{P(+|C) \cdot P(C) + P(+|\neg C) \cdot P(\neg C)}$

$$= \frac{0.9 * 0.01}{0.9 * 0.01 + 0.2 * 0.99} = 0.0434$$

Bayes Rule – graphical representation



$P(A)$

$P(B|A)$
$P(B|\neg A)$

- B – observable but A not observable
- Diagnostic Reasoning: $P(A|B)$ or $P(A|\neg B)$

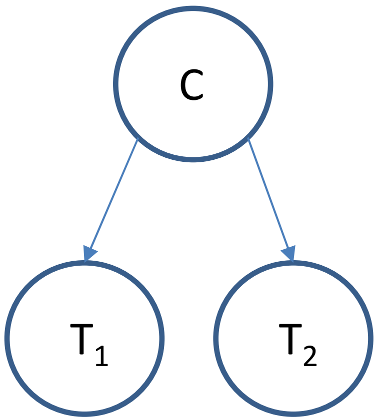
Bayes Rule trick

- $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$, what happens if I don't know the P(B) ?
- $P(\neg A|B) = \frac{P(B|\neg A) \cdot P(\neg A)}{P(B)}$
- The normalizer P(B) is identical
- $P(A|B) + P(\neg A|B) = 1$
- $P'(A|B) = P(B|A) \cdot P(A)$
- $P'(\neg A|B) = P(B|\neg A) \cdot P(\neg A)$
- $P(A|B) = \eta P'(A|B)$
- $P(\neg A|B) = \eta P'(\neg A|B)$
- $\eta = \frac{1}{P'(A|B) + P'(\neg A|B)}$

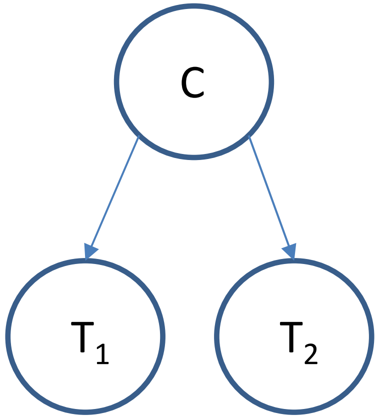
Bayes Rules – cancer example

- $P(C) = 0.01 \,, P(\neg C) = 0.99$
- $P(+|C) = 0.9 \;,\; P(-|C) = 0.1$
- $P(+|\neg C) = 0.2 \;,\; P(-|\neg C) = 0.8$
- $P(C|T_1 = +, T_2 = +) = P(C|++) = ?$

	Prior	+	+	P’	P(C|++)
C	0.01	0.9	0.9	0.0081	0.1698
¬C	0.99	0.2	0.2	0.0396	
				0.0477	

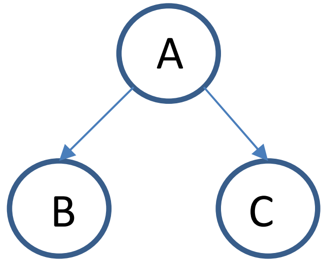


Conditional Independence



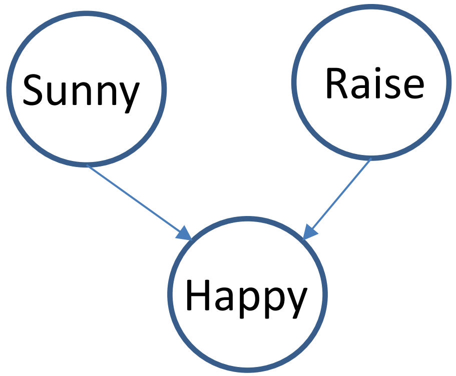
- T_1 and T_2 are conditionally independent
- Knowing anything about T_1 would not help us make a statement about T_2
- $P(T_2|C,T_1) = P(T_2|C)$
- The independence only holds true if we know C (see the diagram)

Conditional Independence



- Given A, $B \perp C$ are independent , written as $B \perp C \mid A$
- B and C are not independent if we don't know A
- $P(T_2 = + | T_1 = +)$=?
- $P(T_2 = + | T_1 = +) = P(+_2 | +_1, C)P(C | +_1) + P(+_2 | +_1, \neg C)P(\neg C | +_1) = P(+_2 | C)P(C | +_1) + P(+_2 | \neg C)P(\neg C | +_1)$=0.9*0.043+0.2*0.957=0.2301

Different type of Bayes Network



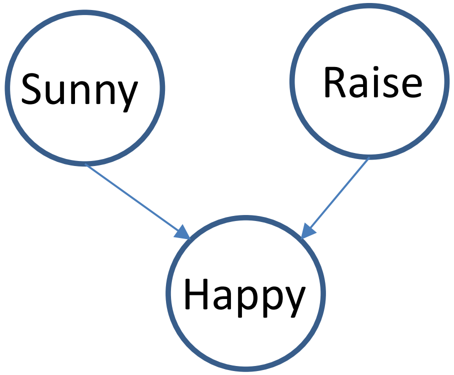
- P(S)=0.7 P(R)=0.01

- P(H|S,R)=1
- P(H| ¬S,R)=0.9
- P(H|S, ¬ R)=0.7
- P(H| ¬S, ¬ R)=0.1

- P(R|S)=?
- P(R|S)=P(R)=0.01

- P(R|H,S)=?
- $P(R|H,S) = \frac{P(H|R,S)\cdot P(R|S)}{P(H|S)} = \frac{P(H|R,S)\cdot P(R)}{P(H|S,R)\cdot P(R)+P(H|S,\neg R)\cdot P(\neg R)} = 0.0142$
- $P(R|H) = \frac{P(H|R)\cdot P(R)}{P(H)} = \frac{[P(H|R,S)\cdot P(S)+P(H|R,\neg S)\cdot P(\neg S)]\cdot P(R)}{P(H|R,S)\cdot P(R,S)+P(H|\neg R,S)\cdot P(\neg R,S)+P(H|R,\neg S)\cdot P(R,\neg S)+P(H|\neg R,\neg S)\cdot P(\neg R,\neg S)} = 0.0185,\quad P(R,S) = P(R)\cdot P(S), \ldots$
- $P(R|H,\neg S) = \frac{P(H|R,\neg S)\cdot P(R|\neg S)}{P(H|\neg S)} = \frac{P(H|R,\neg S)\cdot P(R)}{P(H|R,\neg S)\cdot P(R)+P(H|\neg R,\neg S)\cdot P(\neg R)} = 0.0833$

Different type of Bayes Network

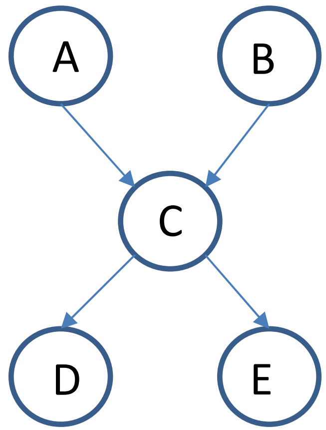


- $P(R|H,S) = 0.0142$
- $P(R|S) = P(R) = 0.01$
- $P(R|H,\neg S) = 0.0833$

- $R \perp S$ if we don't know anything about H
- $not(R \perp S)$ if we know about H
- Knowledge of H makes 2 variables that previously were independent, non-independent
- Independence does not imply conditional independence

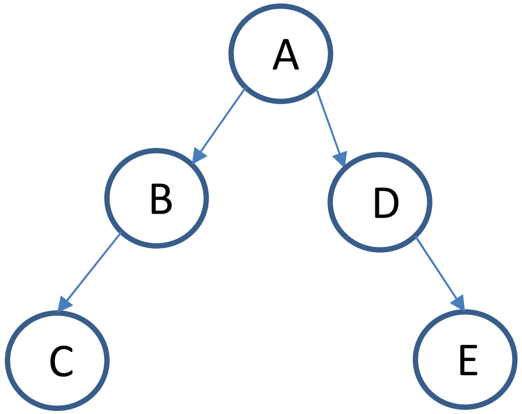
Bayes Networks

- Bayes networks define probability distribution over graphs or random variables

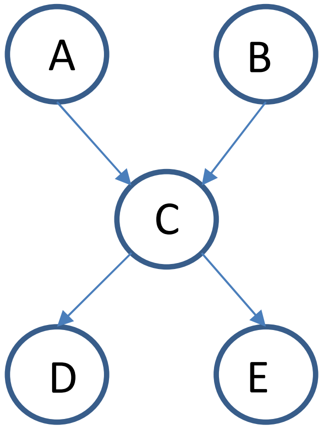


- Bayes network defines the distribution over those 5 random variables
- The Bayes network is defined by probability distributions that are inherent to each individual node
- For node A and B, we have a distribution P(A) and P(B), A and B have no incoming arcs
- For node C we have a distribution P(C|A,B)
- For node D and E we have P(D|C) and P(E|C)
- P(A,B,C,D,E)=P(A)P(B)P(C|A,B)P(D|C)P(E|C)
- Whereas the joint distribution over any 5 variables requires 2^5-1 =31, but the Bayes network requires only 10 variables (1+1+4+2+2=10)
- QUIZ: How many parameters do we need to specify the network defined on the Slide 2 (naïve joint over 16 variables is 2^{16}-1?
- QUIZ ANSWER: 47

D-SEPARATION



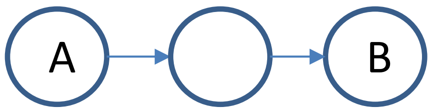
- $C \perp A$ (False)
- $C \perp A \mid B$ (True)
- $C \perp D$ (False)
- $C \perp D \mid A$ (True)
- $E \perp C \mid D$ (True)

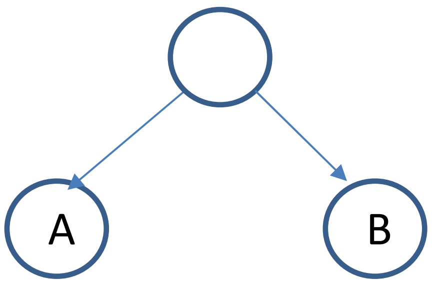


- $A \perp E$ (False)
- $A \perp E \mid B$ (False)
- $A \perp E \mid C$ (True)
- $A \perp B$ (True)
- $A \perp B \mid C$ (False)

D-SEPARATION

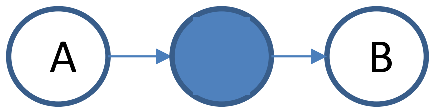
Active triplets
(render variables dependent $\mathrm{not}(A \perp B)$)



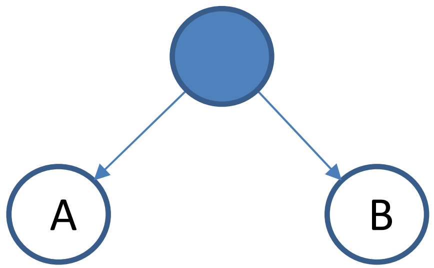


Inactive triplets
(render variables independent $A \perp B$)

Known variable

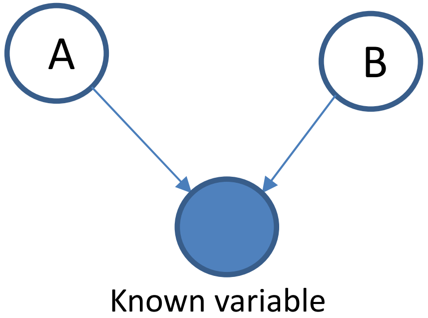


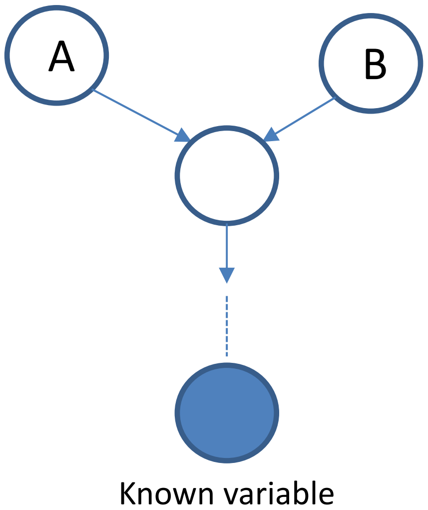
Known variable



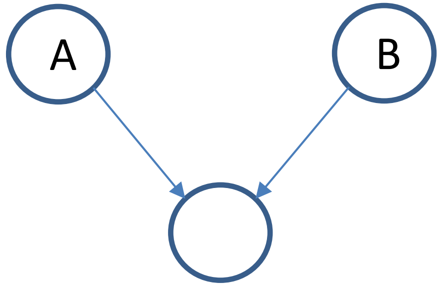
D-SEPARATION

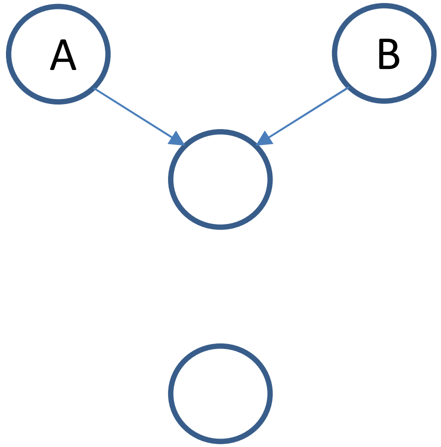
Active triplets
(render variables dependent $\mathrm{not}(A \perp B)$)



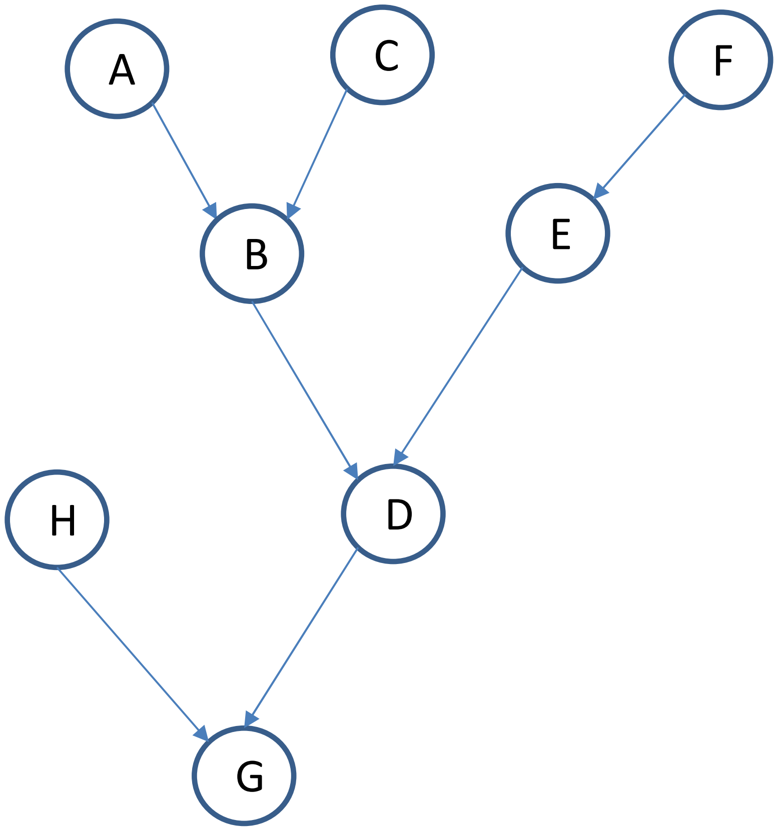


Inactive triplets
(render variables independent $A \perp B$)





D-SEPARATION



- $F \perp A$ (True)
- $F \perp A \mid D$ (False)
- $F \perp A \mid G$ (False)
- $F \perp A \mid H$ (True)