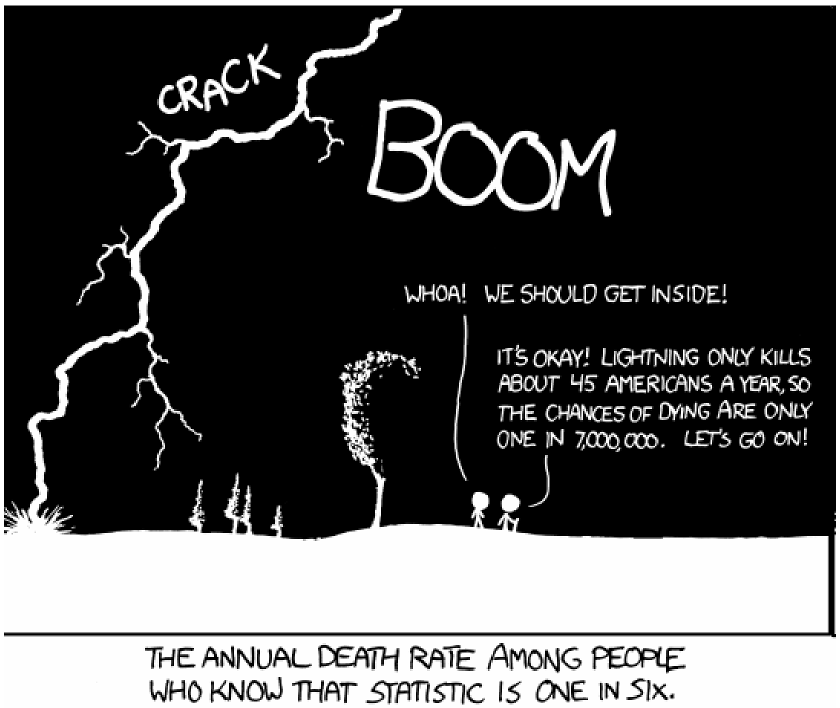
CSE4403 3.0/CSE6602E - Soft Computing
Winter 2011



Lecture 13

Discussion on Bayesian Networks

Guest lecturer: Xiangdong An
xan@cs.yorku.ca

Bayesian networks

- A Bayesian network is a triplet $(V, G, \mathcal{P})$. V is a set of variables, G is a connected DAG whose nodes correspond one-to-one to members of V such that each variable is conditionally independent of its non-descendants given its parents.

  Denote the parents of $v \in V$ in G by $\pi(v)$. $\mathcal{P}$ is a set of probability distributions:

$$\mathcal{P} = \{ P(v | \pi(v)) \mid v \in V \}.$$

Knowledge representation and inference

- Bayesian networks (BNs) are a graphical model for uncertain knowledge representation
  - Can be constructed based on expert knowledge
  - Can be learnt from data
- BNs are a graphical model for reasoning about the state of the problem domains
  - An interpretation to the world, e.g. the posterior probabilities of some variable given evidence
  - To support automated decision making

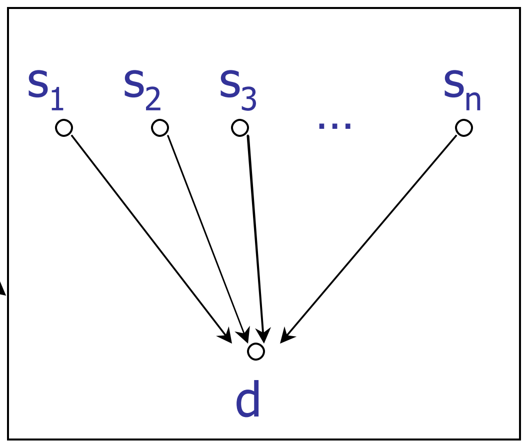
Qualitative structure and quantitative distributions

- A BN consists of two parts, structure and parameters
  - The graphical structure encodes conditional dependencies
    - Qualitatively
  - The probability distribution parameters specify the strength of such dependencies
    - Quantitatively
- This allows us to first focus on qualitative structure and then quantitative strength of dependencies in construction

Causal relationship makes BNs sparse

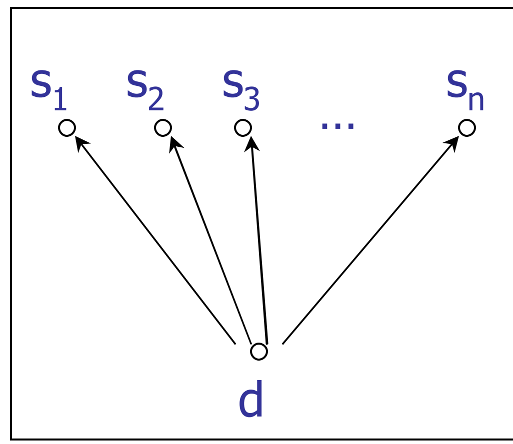
- BNs constructed based on causal (natural) relationship tends to be sparse

If you specify a probability distribution table, it has size at least 2^n



Following non-causal relationship between disease *d* and symptoms $s_1, s_2, \ldots, s_n$

n small tables to specify, which is linear



Following causal relationship from disease *d* to symptoms $s_1, s_2, \ldots, s_n$

Conditional independent

- Let X, Y, and Z be disjoint sets of variables. X and Y are *conditionally independent* give Z, denoted $I(X, Z, Y)$, iff for every $x \in D_X$, $y \in D_Y$, $z \in D_Z$ such that $P(y, z) > 0$, the following holds:

$$P(x|y, z) = P(x|z)$$

Degenerate when Z is ∅

When Z is empty, X and Y are marginally independent, denoted by $I(X, \varnothing, Y)$

$$P(x|y) = P(x)$$

Conditional independent example

$$P(X|Y) = P(X)$$

Y	X	P(X|Y)
0	0	0.3
0	1	0.7
1	0	0.3
1	1	0.7

This pattern repeats

If $P(X|Y) = P(X)$,
whether $P(Y|X) = P(Y)$?

Conditional independent example

$P(X|Y,Z) = P(X|Z)$

Y	Z	X	P(X|Y,Z)
0	0	0	0.1
0	0	1	0.9
0	1	0	0.8
0	1	1	0.2
1	0	0	0.1
1	0	1	0.9
1	1	0	0.8
1	1	1	0.2

This pattern repeats

Decomposition over structures

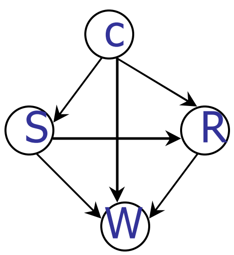
- A Bayesian network is a triplet $(V, G, \mathcal{P})$. V is a set of variables, G is a connected DAG whose nodes correspond one-to-one to members of V such that each variable is conditionally independent of its non-descendants given its parents.

  Denote the parents of $v \in V$ in G by $\pi(v)$. $\mathcal{P}$ is a set of probability distributions:

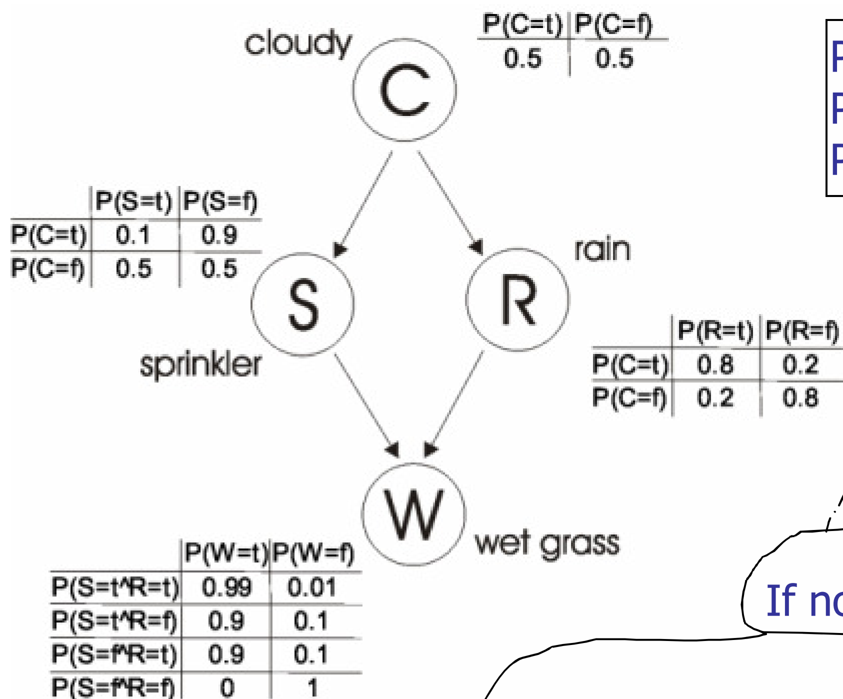
  $\mathcal{P} = \{ P(v|\pi(v)) \mid v \in V\}$.

- By chain rule: $P(V) = \prod_{v \in V} P(v \mid \pi(v))$

Decomposition over structures



Equivalent structure



$P(S|C,R,W)P(C,R,W)$,
$P(R|S,C,W)P(S,C,W)$, or
$P(C|S,R,W)P(S,R,W)$

or

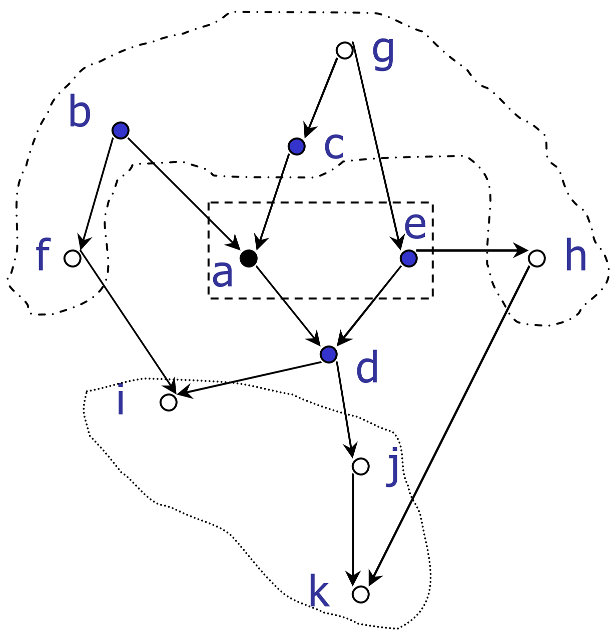
If no structure, i.e. solid

$$
\begin{aligned}
P(C, S, R, W) &= P(W|S,R,C)P(S,R,C) = P(W|S,R,C)P(S,R|C)P(C) \\
&= P(W|S,R,C)P(S|R,C)P(R|C)P(C) \\
&= P(W|S,R,C)P(S|C)P(R|C)P(C) \\
&= P(W|S,R)P(S|C)P(R|C)P(C)
\end{aligned}
$$

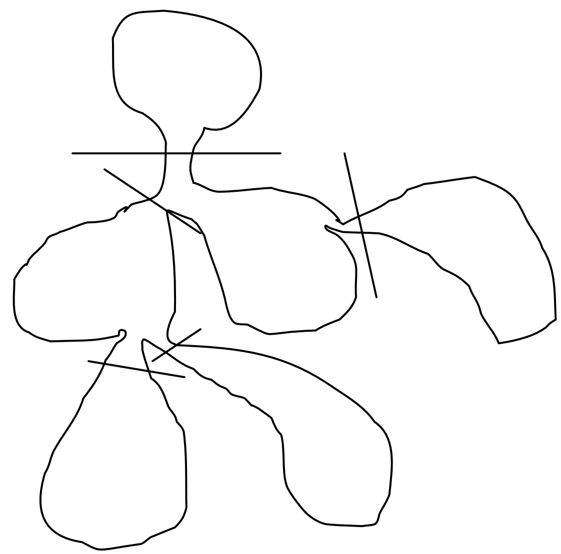
Decompose ↓

Chain rule ↑

Cut a structure through



Not through



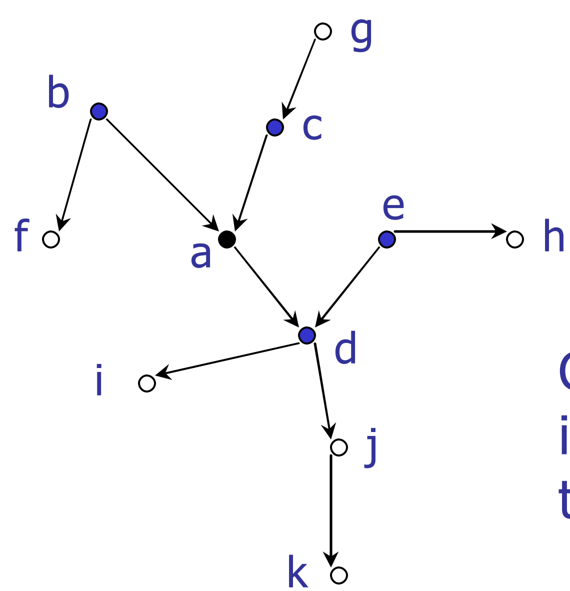
Through

By cut through, we divide and conquer

Markov blanket

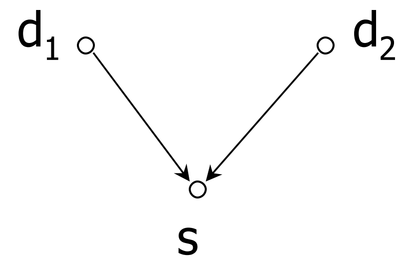
- A Markov blanket of a node includes its parents, children, and children's parents. Given a Markov blanket, the node is independent of all other nodes.

Given a Markov blanket, a graph can be cut through



Given b, c, d, e, a is independent of the *rest* of the structure f, g, h, i, j, k

Why children's parents - v structure

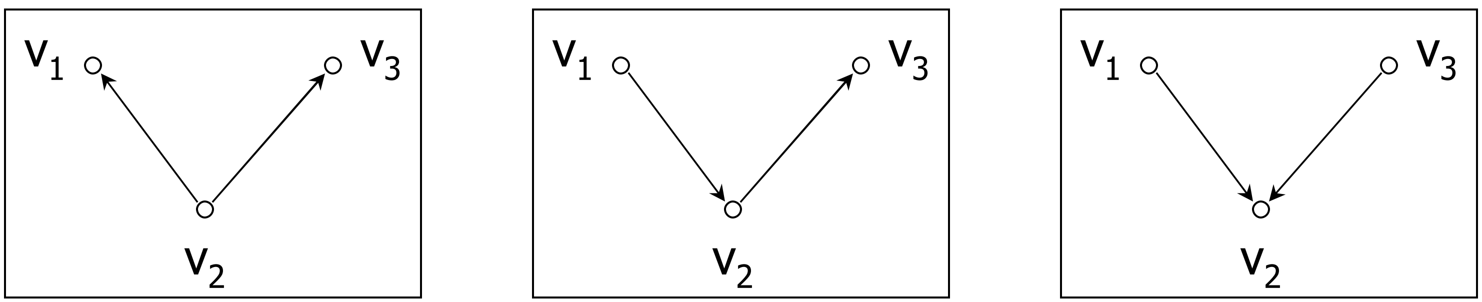


Two diseases d1 and d2 can both cause symptom s. Before we know a patient has symptom s, d1 and d2 could be independent, e.g. headache or fever could be caused by many independent diseases. How if we know a patient has symptom s?

This is also the reason why deductive reasoning can generally only be done in one direction

When two arcs meet

- In a directed graph, when two arcs meet in a path, the shared node can be in one of the three possible cases: tail-to-tail, head-to-tail or head-to-head, as the node v_2 shown below.



Path open or closed

- Let X, Y and Z be disjoint subsets of nodes (vertices) in a DAG G.

  A path ρ between nodes $x \in X$ and $y \in Y$ is rendered *closed* by Z whenever one of the two conditions is true:

  I. There exists $z \in Z$ that is either tail-to-tail or head-to-tail on ρ

  II. There exists a node w that is head-to-head on ρ and neither w nor any descendant of w is in Z. If both conditions are false, then ρ is rendered *open* by Z

Graphical separation

- If every path between x and y is closed by Z, then x and y are said to be separated by Z. X and Y are said to be *separated* by Z if every pair $x \in X$ and $y \in Y$ are separated by Z. We use the notation $< X|Z|Y >_G$ to denote that X and Y are separated by Z in graph G

Graphical separation & independence

- A DAG is an *I-map* (or called *independence map*) of a probability distribution $P(V)$ over a set of variables V , if there is a one-to-one correspondence between nodes in G and variables in V and for every disjoint subsets X, Y and Z, we have

$$< X|Z|Y >_G \Rightarrow I(X, Z, Y)_M$$

- A graph is a *minimal I-map* if all links in it are necessary for it to remain an I-map
  - When an I-map is minimal, there would be no nonwarranted dependency claims
  - Therefore, it is sparser and computation is more efficient

Relationship between structure and distribution

- In a BN, its structure should be an I-map (better if minimal I-map) of its $P(V)$

$$< X|Z|Y >_G \Rightarrow I(X, Z, Y)_M$$

A complete graph is an I-map of any distribution

If X and Y are separated by Z in G, then X and Y should be independent given Z in $P(V)$.

Equivalent

Reversely, if X, Y are dependent in $P(V)$, then X and Y should be dependent in its structure G

Perfect map

- If it is only an I-map, there is no guarantee that independencies in P(V) will have corresponding separation in the structure
- Similarly, if it is only an I-map, there is no guarantee that non-separations in the structure indicates dependencies in P(V)
- To make both guaranteed, the structure should be a perfect-map of P(V), the structure should also be a *D-map* of P(V), but may not always be possible