Bayesian Networks

Discussion:
Practice with Bayesian Networks

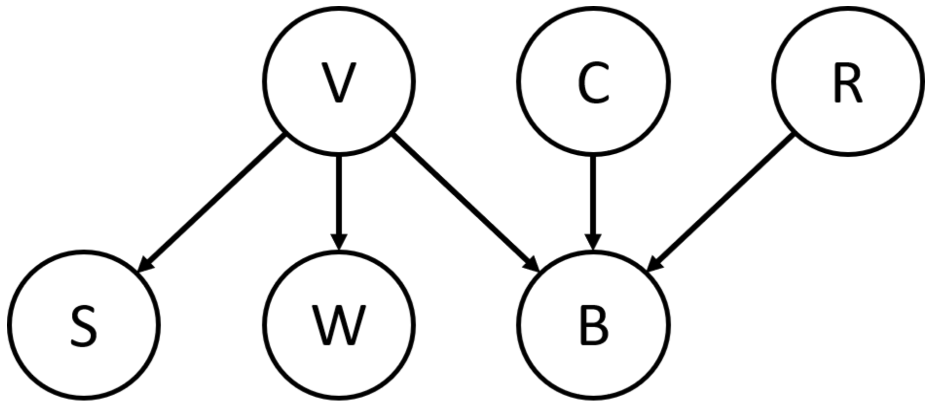
ECE/CS 498 DS

University of Illinois

Agenda

- Joint probability distribution
- Information vs causal flow
- Inferencing with Bayesian Networks
- Calculating parameters required for computations

ICA 2 Example



- Virus (V) ∈ {Yes, No}
- Spam (S) ∈ {Yes, No}
- Warning from Antivirus Software (W) ∈ {Yes, No}
- Configuration (C) ∈ {High, Medium, Low}
- Resource Utilization (R) ∈ {High, Low}
- Speed of Bayes Engine (B) ∈{Fast, Slow}

Joint Probabilities

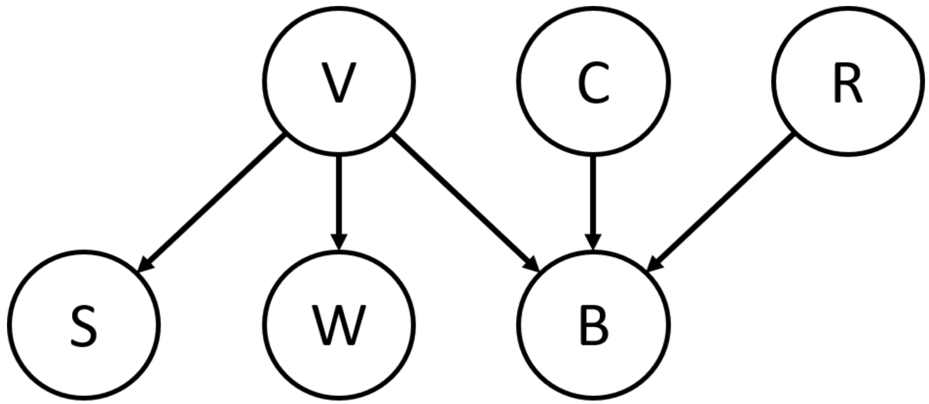
Joint Probability Distribution

- In general, a joint probability distribution describes the probability of multiple events happening at the same time
  - This is very useful when trying to predict outcomes in systems with multiple variables
  - Conditional probabilities can be derived using joint probabilities
- To specify arbitrary joint distribution, we need *many* parameters
  - **Parameter** = known probability value from existing data
  - # parameters required
    = (# unique combinations of variable values) -1
  - "-1" comes from the fact that all joint probability values sum to 1
    - If we enumerate *n* combinations of variable values, and we know the first *n-1*, then

$$parameter_n = 1 - \sum_{k=1}^{n-1} parameter_k$$

  - Sometimes, these are just found empirically with data

Joint Probability Distribution

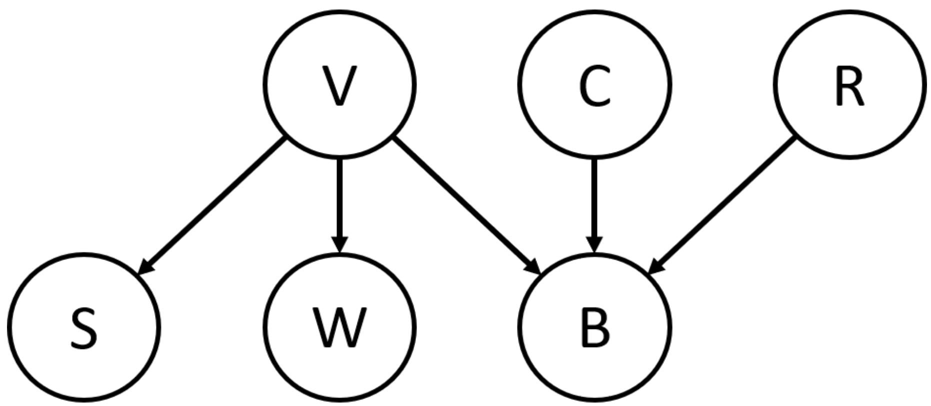


- Example: calculate number of parameters (without Bayesian Network) for joint distribution $P(V, S, W, C, R, B)$
  - $\{V, S, W, R, B\}$ can all take on 2 values, while C takes on 3 values
  - Thus, to specify $P(V, S, W, C, R, B)$, we need (2^5*3) -1 = 95 parameters!

Information vs Causal Flow

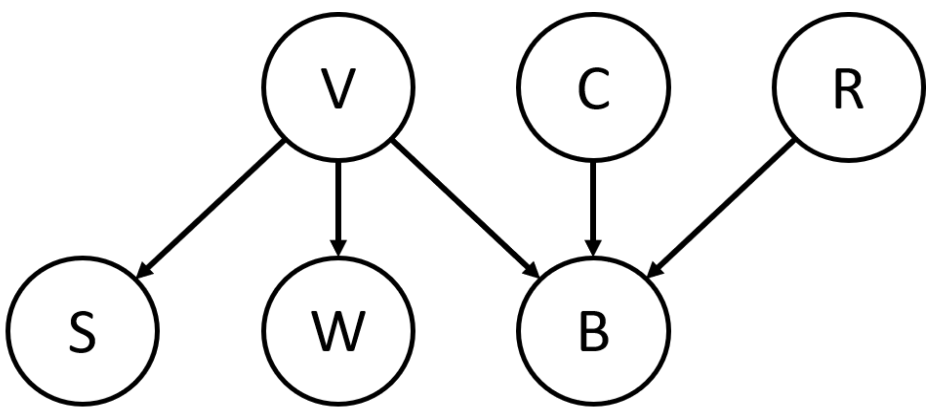
Causal Flow

- So far, we have just studied causal flow in Bayesian Networks
- **<u>Causal flow</u>** defines the path of causality between variables in a Bayesian Network
  - For example, a directed edge from V to S indicates that the presence of a virus (V) ***causes*** some change in behavior of the spam messages (S)



Information Flow

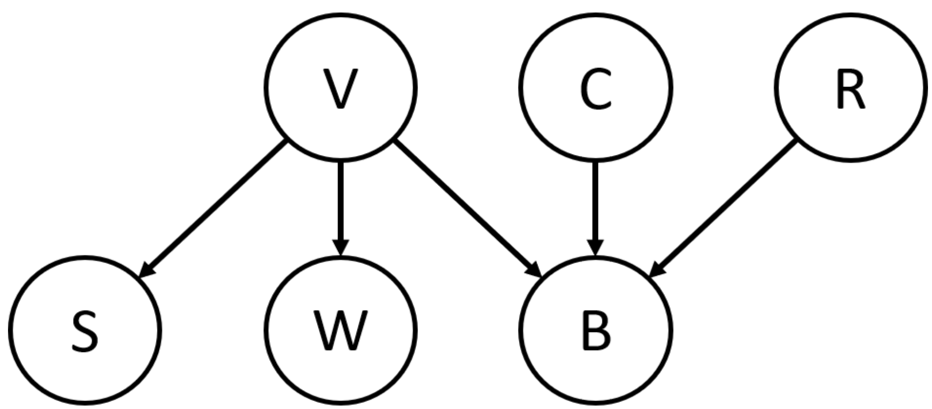
- In addition to causal flow, we can also track **information flow**, which describes how variables provide information about one another
- For example,
  - Does knowing something about V change your belief about R?
  - Given B, does knowing something about W change your belief about C?



Trail Review

- A **trail** is a sequence of vertices/edges in a graph where no edge is repeated
  - e.g. $W \leftarrow V \rightarrow B \leftarrow R$ and $S \leftarrow V \rightarrow B$ are trails
  - Arrows must match directed edges in graph

- A **v-structure** is a specific trail with three nodes consisting of two parents and a common child
  - e.g. $V \rightarrow B \leftarrow R$ and $C \rightarrow B \leftarrow R$ are two v-structures
  - **Can you find the last v-structure?**



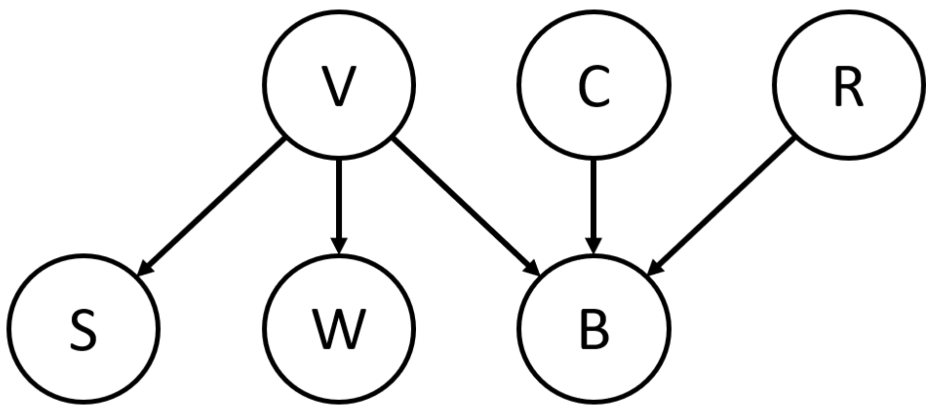
Flow of Probabilistic Influence

- Suppose that '–' represents an arbitrary directional arrow…
- A given trail $\boldsymbol{X_1} - \ldots - \boldsymbol{X_n}$ is said to be ***active*** if there is influence between $\boldsymbol{X_1}$ and $\boldsymbol{X_n}$
  - Only 1 active trail is needed for there to be probabilistic influence between $\boldsymbol{X_1}$ and $\boldsymbol{X_n}$
- A trail $\boldsymbol{X_1} - \ldots - \boldsymbol{X_n}$ is **active** if:
  - It has no v-structures
- A trail $\boldsymbol{X_1} - \ldots - \boldsymbol{X_n}$ is **active given** $\boldsymbol{Z} = \{\boldsymbol{Z_1}, \boldsymbol{Z_2}, \ldots, \boldsymbol{Z_l}\}$ if:
  - For every v-structure $\boldsymbol{X_{i-1}} \rightarrow \boldsymbol{X_i} \leftarrow \boldsymbol{X_{i+1}}$, the collider node $\boldsymbol{X_i}$ or any of its descendants are in $\boldsymbol{Z}$
  - No non-collider node $\boldsymbol{X_i}$ on the trail is in $\boldsymbol{Z}$

Reference: Koller, Probabilistic Graphical Models
https://www.youtube.com/watch?v=JSrNWurmyLU

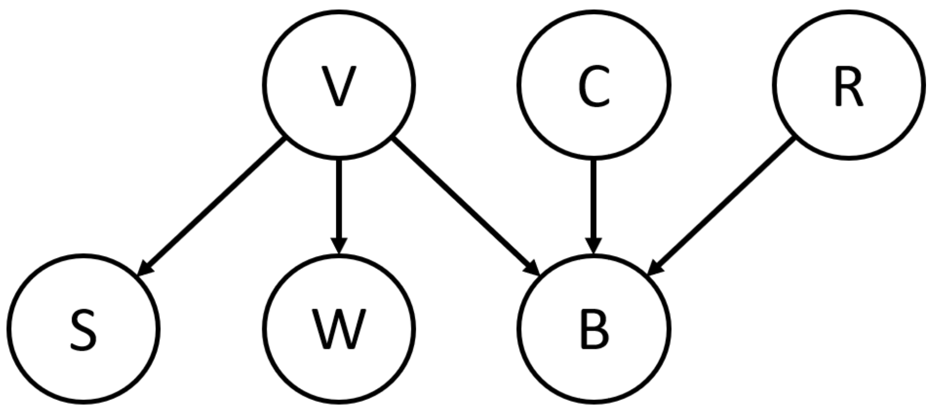
Flow of Probabilistic Influence

- Is $S \perp B$?
  - The only trail between S and B is $S \rightarrow V \rightarrow B$. There are no v-structures on this trial. Thus, this is an active trail, and the independence relationship does not hold (i.e. there is probabilistic influence between S and B).



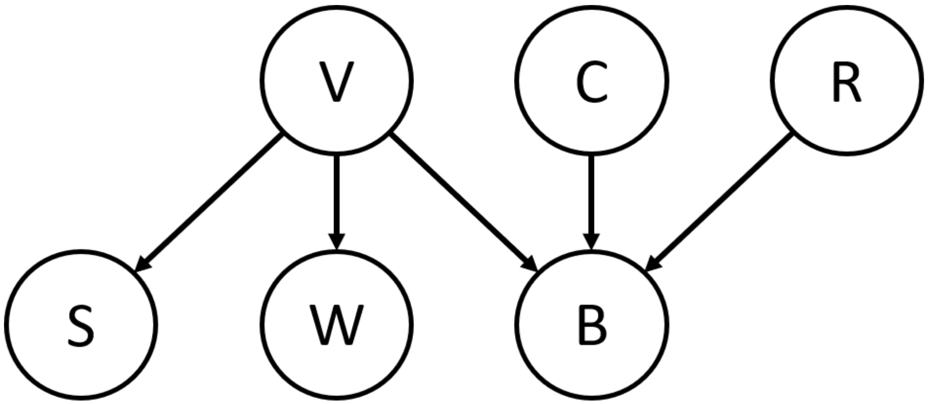
Flow of Probabilistic Influence

- Is $V \perp R$?
  - The only trail between V and R is $V \rightarrow B \leftarrow R$. Since B is a collider node in a v-structure, this trail isn't active. Therefore, the independence relationship holds.



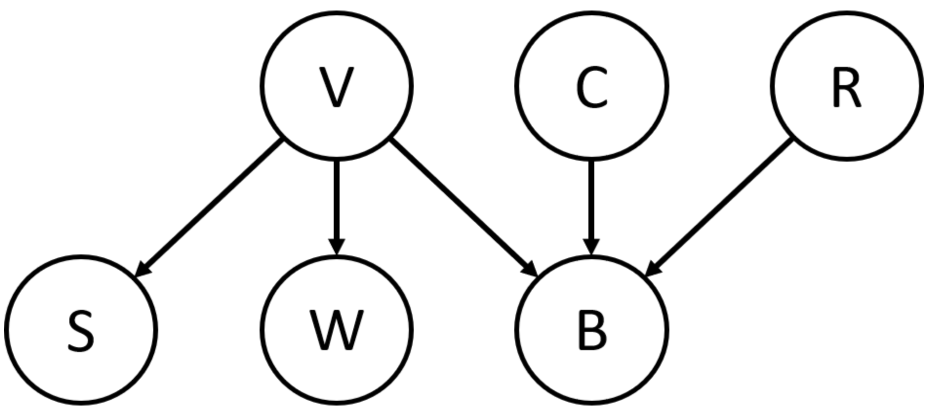
Flow of Probabilistic Influence

- Is $V \perp R | B$?
  - The only trail between V and R is $V \rightarrow B \leftarrow R$. Since B is a collider node in a v-structure, and since B is observed, the trail is active. Therefore, the independence relationship does not hold (i.e. there is probabilistic influence between V and R).



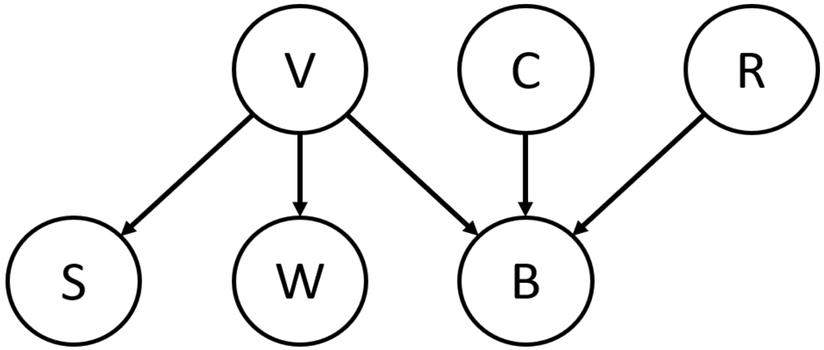
More Practice

- Is $W \perp C|S$?
- Is $V \perp B$?
- Is $S \perp R|B$?



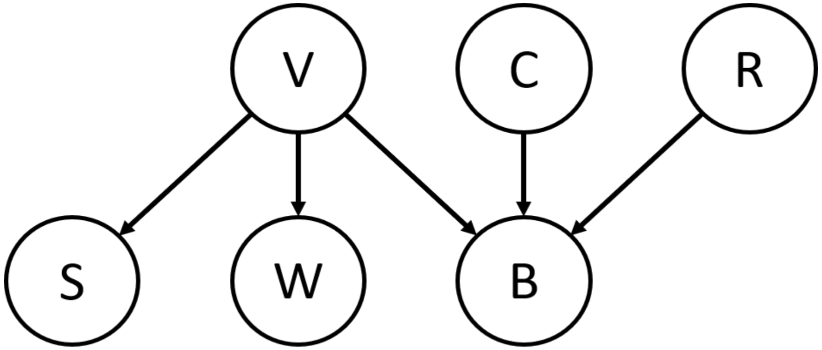
Inference Example

Inference Task



- Given the following information, you need to use the provided CPTs to decide if there is a virus in the system
  - Configuration (C) = High
  - Resource Utilization (R) = High
  - Spam Message (S) = Yes
  - Anti-virus Warning (W) = Yes
  - Bayes Speed (B) = Slow

Inference Task



- In other words,

Is P(V=Yes | C=High, R=High, S=Yes, W=No, B=Slow) > P(V=No | C=High, R=High, S=Yes, W=No, B=Slow)?

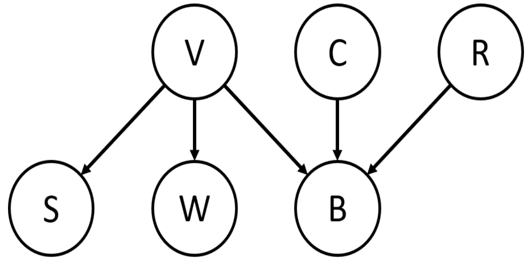
  - Alternatively, is Is P(V=Yes | C=High, R=High, S=Yes, W=No, B=Slow) > 0.5?
  - We generally don't use the latter approach as it requires computing the normalizing factor for the posterior…

Conditional Probability Tables

P(V,S,W,B,C,R) = P(S|V)P(W|V)P(B|V,C,R)P(C)P(V)P(R)

V (Virus)	C (Configuration)	R (Utilization)	P(B=Slow|V,C,R)
Yes	Low	Low	0.8
Yes	Low	High	0.9
Yes	Medium	Low	0.5
Yes	Medium	High	0.75
Yes	High	Low	0.4
Yes	High	High	0.6
No	Low	Low	0.25
No	Low	High	0.4
No	Medium	Low	0.1
No	Medium	High	0.35
No	High	Low	0.05
No	High	High	0.2

Configuration (C) = High
Resource Utilization (R) = High
Spam Message (S) = Yes
Anti-virus Warning (W) = Yes
Bayes Speed (B) = Slow



V (Virus)	P(W=Yes|V)
Yes	0.7
No	0.2

V (Virus)	P(S=Yes|V)
Yes	0.6
No	0.2

P(V=Yes) = 0.2   P(C=High) = 0.25   P(R=High) = 0.4

$$P(V = Yes \mid C = High, R = High, S = Yes, W = No)$$
$$\propto P(S = Yes|V = Yes) * P(W = No|V = Yes)$$
$$* P(B = Slow|V = Yes, C = High, R = High) * P(C = High)$$
$$* P(R = High) * P(V = Yes)$$
$$= 0.6 * 0.3 * 0.6 * 0.25 * 0.4 * 0.2 = \mathbf{0.00216}$$

$$P(V = No \mid C = High, R = High, S = Yes, W = No)$$
$$\propto P(S = Yes|V = No) * P(W = No|V = No)$$
$$* P(B = Slow|V = No, C = High, R = High) * P(C = High)$$
$$* P(R = High) * P(V = No)$$
$$= 0.2 * 0.8 * 0.2 * 0.25 * 0.4 * 0.8 = \mathbf{0.00256}$$

Thus, using MAP rule, we decide that there is no virus in the system.

Counting Required Parameters

Counting Parameters

- After factorizing the joint distribution per the Bayesian Network, we arrive at

$$P(V,S,W,B,C,R) = P(S|V)P(W|V)P(B|V,C,R)P(C)P(V)P(R)$$

- **How do we calculate the number of required parameters for this factorization?**

Counting Parameters

$$P(V,S,W,B,C,R) = P(S|V)P(W|V)P(B|V,C,R)P(C)P(V)P(R)$$

Consider $P(R)$

$P(R = Low)$	$P(R = High)$
k	$1 - k$

- We only need 1 parameter in order to fully determine $P(R)$

Counting Parameters

$$P(V,S,W,B,C,R) = P(S|V)P(W|V)P(B|V,C,R)P(C)P(V)P(R)$$

- Consider $P(W|V)$

Virus	$P(W = Yes|V)$	$P(W = No|V)$
V=Yes	k_1	$1 - k_1$
V=No	k_2	$1 - k_2$

- We only need 2 parameters in order to fully determine $P(W|V)$

Counting Parameters

$P(V,S,W,B,C,R) = P(S|V)P(W|V)P(B|V,C,R)P(C)P(V)P(R)$

- Now, how many total parameters are required?

- $P(C)$: 2, $P(R)$: 1, $P(V)$: 1, $P(B|V,C,R)$: 12, $P(W|V)$:2, $P(S|V)$:2
- Total = 2 + 1 + 1 + 12 + 2 + 2 = 20