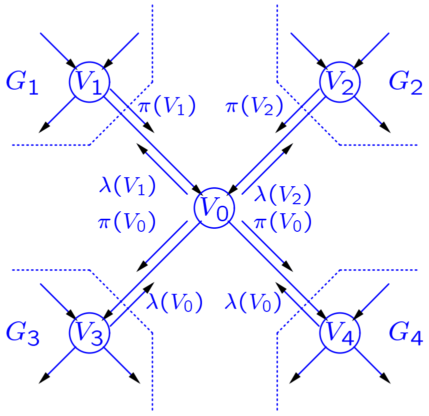
Bayesian Belief Networks III

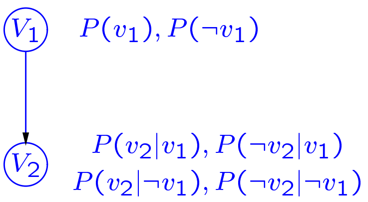
- **Stochastic independence:** basis for compact representation and faster probabilistic inference
- **Probabilistic inference algorithms:**
  - 'Naive' Bayesian inference
  - Pearl's algorithm (Judea Pearl, UCLA)
  - Algorithm by Spiegelhalter & Lauritzen (David Spiegelhalter, Cambridge University; Stefen Lauritzen, University of Aalborg) – implementation version by Finn Jensen
- **Notation:** V_i's denote variables and vertices (nodes) at the same time

Basic Idea of Pearl's Algorithm



- Object-oriented approach: vertices are objects, which have local information and carry out local computations
- Updating of probability distribution by message passing: arcs are communication channels

Probabilistic Inference as Message Passing

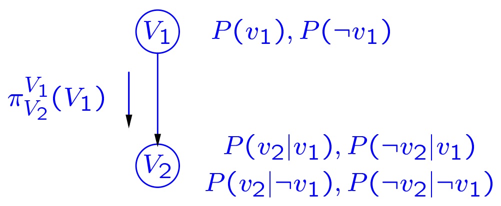


- For vertex V_1: $P(v_1)$ and $P(\neg v_1)$ are known
- For vertex V_2: $P(v_2)$ and $P(\neg v_2)$ are *unknown*; $P(V_2|V_1)$ is known
- It holds that:

$$\begin{aligned} P(v_2) &= P(v_2|v_1)P(v_1) + P(v_2|\neg v_1)P(\neg v_1) \\ P(\neg v_2) &= P(\neg v_2|v_1)P(v_1) + \\ &\quad P(\neg v_2|\neg v_1)P(\neg v_1) \end{aligned}$$

Conclusion: V_2 needs $P(V_1)$, which is sent from V_1 to V_2: $\pi_{V_2}^{V_1}(V_1)$

Message Passing: Causal Parameter $\pi_{V_j}^{V_i}$



It holds that:

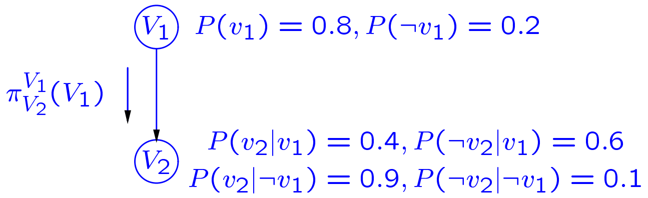
$$\begin{aligned} \pi_{V_2}^{V_1}(v_1) &= P(v_1) \\ \pi_{V_2}^{V_1}(\neg v_1) &= P(\neg v_1) \end{aligned}$$

Local computation in V_2:

$$\begin{aligned} P(v_2) &= P(v_2|v_1)\pi_{V_2}^{V_1}(v_1) + P(v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1) \\ P(\neg v_2) &= P(\neg v_2|v_1)\pi_{V_2}^{V_1}(v_1) + \\ &\quad P(\neg v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1) \end{aligned}$$

$\pi_{V_j}^{V_i}$ is called a causal parameter

Example Causal Parameter $\pi_{V_j}^{V_i}$



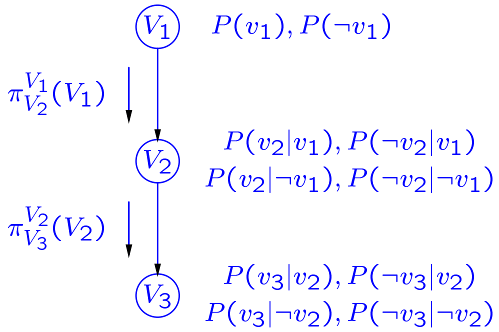
We have:

$$\pi_{V_2}^{V_1}(v_1) = P(v_1) = 0.8$$
$$\pi_{V_2}^{V_1}(\neg v_1) = P(\neg v_1) = 0.2$$

Local computation in V_2:

$$\begin{aligned} P(v_2) &= P(v_2|v_1)\pi_{V_2}^{V_1}(v_1) + P(v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1) \\ &= 0.4 \times 0.8 + 0.8 \times 0.2 = 0.5 \\ P(\neg v_2) &= P(\neg v_2|v_1)\pi_{V_2}^{V_1}(v_1) + \\ &\quad P(\neg v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1) \\ &= 0.6 \times 0.8 + 0.1 \times 0.2 = 0.5 \end{aligned}$$

Message Passing: Three Vertices



It holds that:

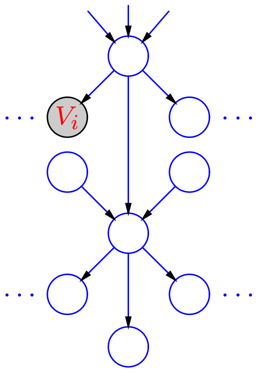
$$\begin{aligned} \pi_{V_3}^{V_2}(v_2) &= P(v_2) \\ \pi_{V_3}^{V_2}(\neg v_2) &= P(\neg v_2) \end{aligned}$$

Local computation in V_3:

$$\begin{aligned} P(v_3) &= P(v_3|v_2)\pi_{V_3}^{V_2}(v_2) + P(v_3|\neg v_2)\pi_{V_3}^{V_2}(\neg v_2) \\ P(\neg v_3) &= P(\neg v_3|v_2)\pi_{V_3}^{V_2}(v_2) + \\ &\quad P(\neg v_3|\neg v_2)\pi_{V_3}^{V_2}(\neg v_2) \end{aligned}$$

Evidence Propagation

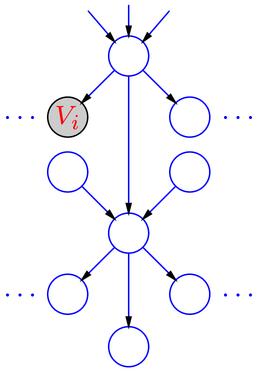
Let $\mathcal{B} = (G, P)$ be a Bayesian belief network with digraph G and joint probability distribution P



- **Evidence:** assignment of value to variable (= instantiating the variable) – V_i = *true* (= v_i) or V_i = *false* (= $\neg v_i$) for binary variable V_i

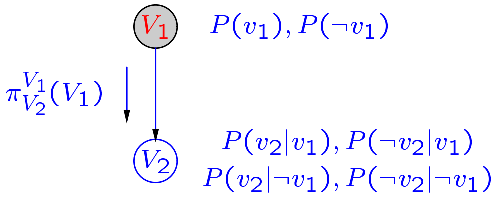
Evidence Propagation

Let $\mathcal{B} = (G, P)$ be a Bayesian belief network



- Consequence: P no longer holds, and must be updated to a new probability distribution P^*. For example: for evidence v_i it holds that $P^*(v_i) = 1$, and thus $P^*(\neg v_i) = 0$, whereas originally we could have had $P(v_i) = 0.3$ and $P(\neg v_i) = 0.7$
- Entire Bayesian belief network must be updated

Evidence Propagation: Causal Parameter



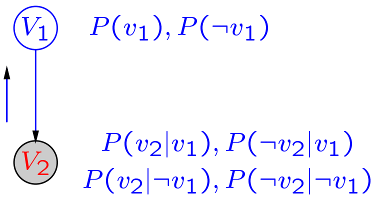
Evidence: assume that $V_1 = \mathit{true}\ (= v_1)$

$$\pi^{V_1}_{V_2}(v_1) = 1$$
$$\pi^{V_1}_{V_2}(\neg v_1) = 0$$

Local computation in V_2:

$$\begin{aligned}
P^*(v_2) &= P(v_2|v_1)\pi^{V_1}_{V_2}(v_1) + P(v_2|\neg v_1)\pi^{V_1}_{V_2}(\neg v_1) \\
&= P(v_2|v_1) \\
P^*(\neg v_2) &= P(\neg v_2|v_1)\pi^{V_1}_{V_2}(v_1) + \\
&\quad P(\neg v_2|\neg v_1)\pi^{V_1}_{V_2}(\neg v_1) \\
&= P(\neg v_2|v_1)
\end{aligned}$$

Evidence Propagation: Diagnostic Parameter



Evidence: assume that $V_2 = \mathit{true}\ (= v_2)$

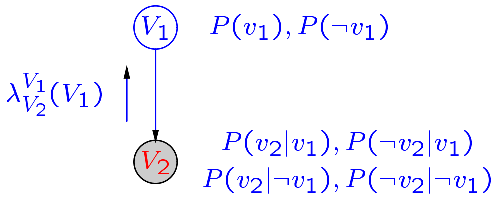
$$P^*(v_2) = 1$$
$$P^*(\neg v_2) = 0$$

Updated probability distribution $P^*(V_1)$ (by local computation in V_1):

$$\begin{aligned}
P^*(v_1) &= P(v_1|v_2) = \frac{P(v_2|v_1)P(v_1)}{P(v_2)} \\
P^*(\neg v_1) &= P(\neg v_1|v_2) = \frac{P(v_2|\neg v_1)P(\neg v_1)}{P(v_2)}
\end{aligned}$$

for which V_1 needs $P(V_2|V_1)$ from V_2

Evidence Propagation: Diagnostic Parameter



Evidence: assume that $V_2 = \mathit{true}\ (= v_2)$

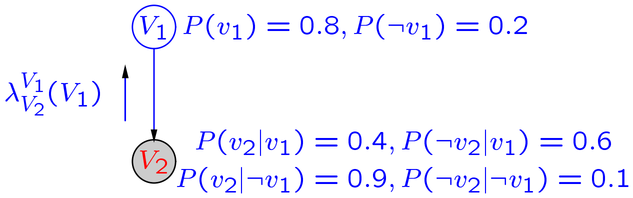
Vertex V_2 sends a message $\lambda^{V_1}_{V_2}(V_1)$ to V_1 so that V_1 is able to compute $P^*(V_1)$

This message is defined as follows:

$$\begin{aligned}
\lambda^{V_1}_{V_2}(v_1) &= P(v_2|v_1) \\
\lambda^{V_1}_{V_2}(\neg v_1) &= P(v_2|\neg v_1)
\end{aligned}$$

$\lambda^{V_1}_{V_2}(V_1)$ is called the **diagnostic parameter**

Example: Diagnostic Parameter

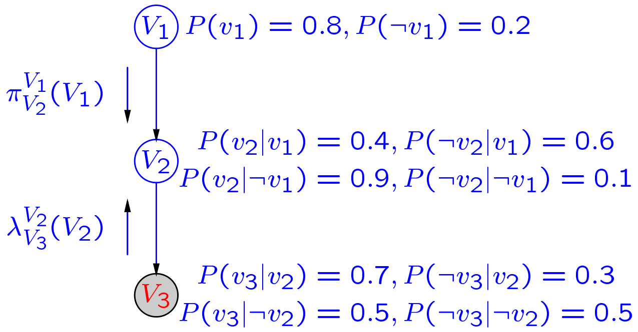


Updated probability distribution $P^*(V_1)$:

$$\begin{aligned}
P^*(v_1) &= \frac{P(v_2|v_1)P(v_1)}{P(v_2)} \\
&= \alpha\, \lambda^{V_1}_{V_2}(v_1)P(v_1) \\
&= \alpha \times 0.4 \times 0.8 = 0.32\alpha \\
P^*(\neg v_1) &= \frac{P(v_2|\neg v_1)P(\neg v_1)}{P(v_2)} \\
&= \alpha\, \lambda^{V_1}_{V_2}(\neg v_1)P(\neg v_1) \\
&= \alpha \times 0.9 \times 0.2 = 0.18\alpha
\end{aligned}$$

with $P^*(v_1) + P^*(\neg v_1) = 1 = 0.32\alpha + 0.18\alpha$
$\Rightarrow \alpha = 1/P(v_2) = 2$

Causal and Diagnostic Parameters Combined

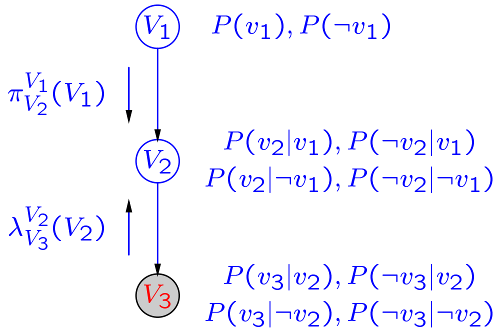


Updated probability distribution $P^*(V_2)$ for *evidence* v_3:

$$\begin{aligned} P^*(v_2) &= \alpha\, \lambda^{V_2}_{V_3}(v_2)[P(v_2|v_1)\pi^{V_1}_{V_2}(v_1) + \\ &\qquad P(v_2|\neg v_1)\pi^{V_1}_{V_2}(\neg v_1)] \\ &= \alpha \times 0.7[0.4 \times 0.8 + 0.9 \times 0.2] = 0.35\alpha \\ P^*(\neg v_2) &= \text{analogous} = 0.25\alpha \; (\Rightarrow \alpha = 1\tfrac{2}{3}) \end{aligned}$$

$\lambda^{V_2}_{V_3}(V_2) = P(V_3|V_2)$, and $\pi^{V_1}_{V_2}(V_1) = P(V_1)$

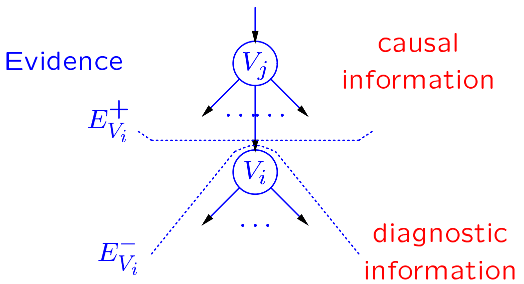
Towards a Generic Formula



Updated probability distribution $P^*(V_2)$ for *evidence* v_3:

$$\begin{aligned} P^*(V_2) &= \alpha\, \lambda^{V_2}_{V_3}(V_2)[P(V_2|v_1)\pi^{V_1}_{V_2}(v_1) + \\ &\qquad P(V_2|\neg v_1)\pi^{v_1}_{V_2}(\neg v_1)] \\ &= \alpha \cdot \text{diagnostic information for } V_2 \cdot \\ &\quad\ \text{causal information for } V_2 \\ &= P(V_2 \mid \text{Evidence}) \end{aligned}$$

Generic Formula: Data Fusion



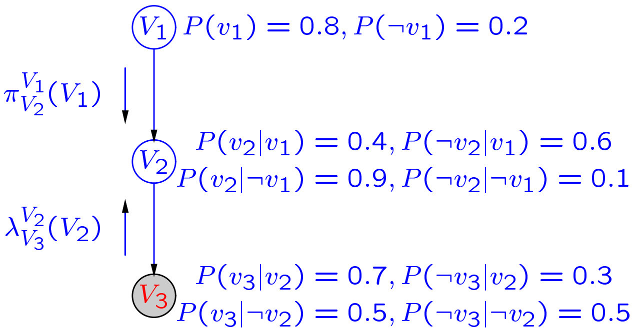
Data fusion:

$$\begin{aligned} P^*(V_i) &= P(V_i \mid E) \\ &= \alpha \cdot \text{diagnostic information for } V_i \cdot \\ &\quad\ \text{causal information for } V_i \\ &= \alpha \cdot \pi(V_i) \cdot \lambda(V_i) \end{aligned}$$

where:

- $E = E^+_{V_i} \cup E^-_{V_i}$: evidence
- α: normalisation constant
- $\pi(V_i)$: *compound* causal parameter
- $\lambda(V_i)$: *compound* diagnostic parameter

Example of Data Fusion



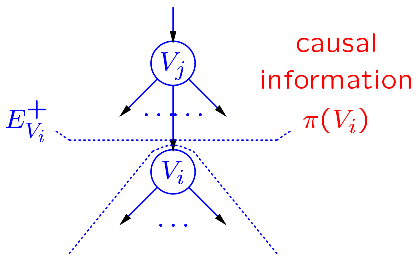
Evidence v_3:

$$P^*(v_2) = P(v_2 \mid v_3) = \alpha \cdot \pi(v_2) \cdot \lambda(v_2) \approx 0.58$$

because:

- $\lambda(v_2) = \lambda^{V_2}_{V_3}(v_2) = 0.7$
- $\pi(v_2) = P(v_2|v_1)\pi^{V_1}_{V_2}(v_1) + P(v_2|\neg v_1)\pi^{V_1}_{V_2}(\neg v_1) = 0.5$
- $\alpha = 1\frac{2}{3}$

Compound Causal Parameters for Trees



Bayesian belief network $\mathcal{B} = (G, P)$, where

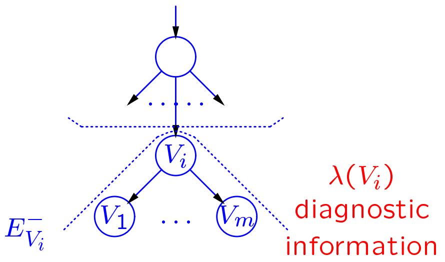
- $G = (V(G), A(G))$ is a directed tree, i.e. exactly one directed path from the root to each vertex $V_i \in V(G)$
- P is a joint probability distribution on $V(G)$

It holds that:

$$\pi(V_i) = \sum_{V_j} P(V_i \mid V_j) \cdot \pi_{V_i}^{V_j}(V_j)$$

with evidence $E^+_{V_i}$ incorporated into $\pi_{V_i}^{V_j}(V_j)$

Compound Diagnostic Parameters for Trees



Bayesian belief network $\mathcal{B} = (G, P)$, where

- $G = (V(G), A(G))$ is a directed tree
- P is the joint probability distribution on $V(G)$

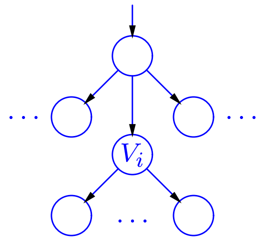
It holds that:

$$\lambda(V_i) = \prod_{j=1}^{m} \lambda_{V_j}^{V_i}(V_i)$$

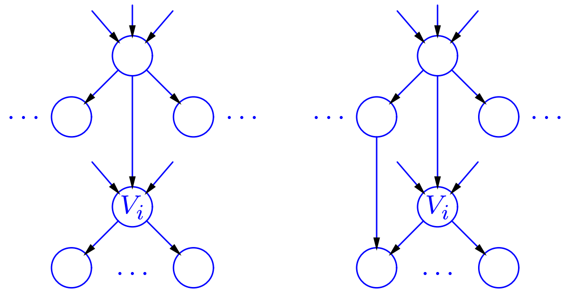
with evidence $E^-_{V_i}$ incorporated into $\lambda_{V_j}^{V_i}(V_i)$

Topology of Bayesian Belief Networks

- **Directed tree:**



- **Singly/multiply-connected digraph:**



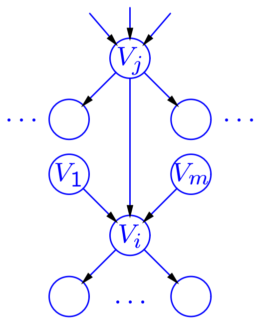
singly-connected multiply-connected

Singly-connected Digraph: Compound Causal Parameter

- **Data fusion:**

$$P^*(V_i) = P(V_i \mid E) = \alpha \cdot \pi(V_i) \cdot \lambda(V_i)$$

- **Compound causal parameter $\pi(V_i)$:**



$$\pi(V_i) = \sum_{\rho(V_i)} P(V_i \mid \rho(V_i)) \cdot \prod_{j=1}^{m} \pi_{V_i}^{V_j}(V_j)$$

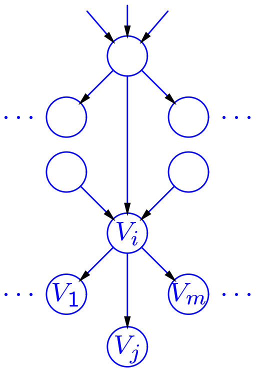
with parents $\rho(V_i) = V_1 \wedge \cdots \wedge V_j \wedge \cdots \wedge V_m$

Singly-connected Digraph: Compound Diagnostic Parameter

- **Data fusion:**

$$P^*(V_i) = P(V_i \mid E) = \alpha \cdot \pi(V_i) \cdot \lambda(V_i)$$

- **Compound diagnostic parameter $\lambda(V_i)$:**

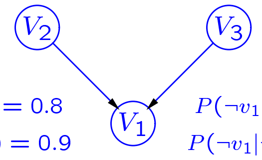


$$\lambda(V_i) = \prod_{j=1}^{m} \lambda_{V_j}^{V_i}(V_i)$$

Example

$P(v_2) = 0.1$
$P(\neg v_2) = 0.9$

$P(v_3) = 0.4$
$P(\neg v_3) = 0.6$



$P(v_1|v_2 \wedge v_3) = 0.8$
$P(v_1|\neg v_2 \wedge v_3) = 0.9$
$P(v_1|v_2 \wedge \neg v_3) = 0.5$
$P(v_1|\neg v_2 \wedge \neg v_3) = 0.6$

$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

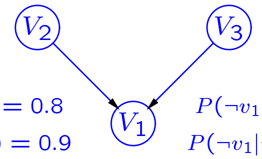
It holds that:

$$\begin{aligned}
P(v_1) &= \alpha \cdot \pi(v_1) \cdot \lambda(v_1) \\
P(\neg v_1) &= \alpha \cdot \pi(\neg v_1) \cdot \lambda(\neg v_1) \\
\pi(v_1) &= P(v_1|v_2 \wedge v_3)\pi_{V_1}^{V_2}(v_2)\pi_{V_1}^{V_3}(v_3) + \\
&\quad P(v_1|\neg v_2 \wedge v_3)\pi_{V_1}^{V_2}(\neg v_2)\pi_{V_1}^{V_3}(v_3) + \\
&\quad P(v_1|v_2 \wedge \neg v_3)\pi_{V_1}^{V_2}(v_2)\pi_{V_1}^{V_3}(\neg v_3) + \\
&\quad P(v_1|\neg v_2 \wedge \neg v_3)\pi_{V_1}^{V_2}(\neg v_2)\pi_{V_1}^{V_3}(\neg v_3) \\
&= 0.8 \times 0.1 \times 0.4 + 0.9 \times 0.9 \times 0.4 + \\
&\quad 0.5 \times 0.1 \times 0.6 + 0.6 \times 0.9 \times 0.6 \\
&= 0.71 \\
\pi(\neg v_1) &= 0.29
\end{aligned}$$

Example

$P(v_2) = 0.1$
$P(\neg v_2) = 0.9$

$P(v_3) = 0.4$
$P(\neg v_3) = 0.6$



$P(v_1|v_2 \wedge v_3) = 0.8$
$P(v_1|\neg v_2 \wedge v_3) = 0.9$
$P(v_1|v_2 \wedge \neg v_3) = 0.5$
$P(v_1|\neg v_2 \wedge \neg v_3) = 0.6$

$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

It holds that:

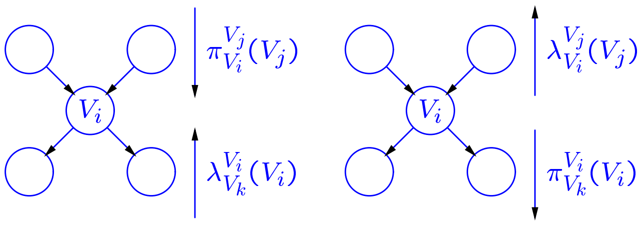
$$\begin{aligned}
P(v_1) &= \alpha \cdot \pi(v_1) \cdot \lambda(v_1) \\
P(\neg v_1) &= \alpha \cdot \pi(\neg v_1) \cdot \lambda(\neg v_1)
\end{aligned}$$

with: $\lambda(v_1) = \prod_{j=1}^{m} \lambda_{V_j}^{V_1}(v_1) = \prod_{\varnothing} = 1$; analogously: $\lambda(\neg v_1) = 1$

Conclusion:

- $P(v_1) = \alpha \cdot 0.71 \cdot 1$; $P(\neg v_1) = \alpha \cdot 0.29 \cdot 1$
- $\Rightarrow \alpha = 1$

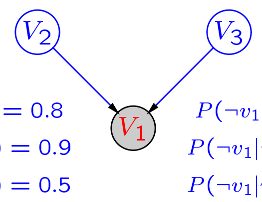
Messages to Children & Parents



Example:

$P(v_2) = 0.1$
$P(\neg v_2) = 0.9$

$P(v_3) = 0.4$
$P(\neg v_3) = 0.6$



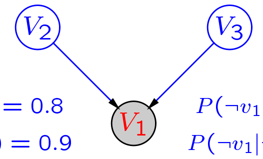
$P(v_1|v_2 \wedge v_3) = 0.8$
$P(v_1|\neg v_2 \wedge v_3) = 0.9$
$P(v_1|v_2 \wedge \neg v_3) = 0.5$
$P(v_1|\neg v_2 \wedge \neg v_3) = 0.6$

$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

- **Evidence:** v_1
- What is $P^*(V_2)$?

Messages to Child & Parent

$P(v_2) = 0.1$
$P(\neg v_2) = 0.9$



$P(v_3) = 0.4$
$P(\neg v_3) = 0.6$

$P(v_1|v_2 \wedge v_3) = 0.8$
$P(v_1|\neg v_2 \wedge v_3) = 0.9$
$P(v_1|v_2 \wedge \neg v_3) = 0.5$
$P(v_1|\neg v_2 \wedge \neg v_3) = 0.6$

$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

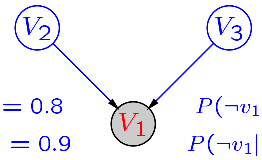
$$P^*(V_2) = \alpha \cdot \pi(V_2) \cdot \lambda(V_2)$$

Lambda message $\lambda_{V_1}^{V_2}(V_2)$ from V_1 to V_2:

$$\begin{aligned}
\lambda_{V_1}^{V_2}(v_2) &= \lambda(v_1) \cdot [P(v_1|v_2 \wedge v_3)\pi_{V_1}^{V_3}(v_3) + \\
&\qquad P(v_1|v_2 \wedge \neg v_3)\pi_{V_1}^{V_3}(\neg v_3)] + \\
&\quad \lambda(\neg v_1) \cdot [P(\neg v_1|v_2 \wedge v_3)\pi_{V_1}^{V_3}(v_3) + \\
&\qquad P(\neg v_1|v_2 \wedge \neg v_3)\pi_{V_1}^{V_3}(\neg v_3)] \\
&= 0.8 \times 0.4 + 0.5 \times 0.6 = 0.62 \\
\lambda_{V_1}^{V_2}(\neg v_2) &= 0.72
\end{aligned}$$

Messages to Child & Parent

$P(v_2) = 0.1$
$P(\neg v_2) = 0.9$



$P(v_3) = 0.4$
$P(\neg v_3) = 0.6$

$P(v_1|v_2 \wedge v_3) = 0.8$
$P(v_1|\neg v_2 \wedge v_3) = 0.9$
$P(v_1|v_2 \wedge \neg v_3) = 0.5$
$P(v_1|\neg v_2 \wedge \neg v_3) = 0.6$

$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

$$P^*(V_2) = \alpha \cdot \pi(V_2) \cdot \lambda(V_2)$$

$$\begin{aligned}
\lambda_{V_1}^{V_2}(v_2) &= 0.62 \\
\lambda_{V_1}^{V_2}(\neg v_2) &= 0.72
\end{aligned}$$

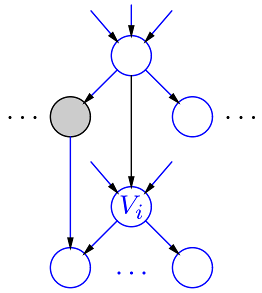
Pi message from $\varnothing$ to V_2: $\pi(v_2) = P(v_2)$, and $\pi(\neg v_2) = P(\neg v_2)$

Result:

$$\begin{aligned}
P^*(v_2) &= \alpha \cdot 0.1 \cdot 0.62 = 0.062\alpha \\
P^*(v_2) &= \alpha \cdot 0.9 \cdot 0.72 = 0.648\alpha
\end{aligned}$$

$\Rightarrow P^*(v_2) \approx 0.087, P^*(\neg v_2) \approx 0.913$

Multiply-connected Digraph: Loop Cutset Conditioning



- Cut every loop (cycle in undirected underlying graph)
- Instantiate the variables with every value at which the loops were cut
- Compute the a posteriori probability distributions
- Resulting probability distributions yield requested distribution