Inference in Bayesian Networks

Pearl's algorithm

The focus of today and next week ...

Main inference problem in graphical models: determine the *marginal probability distribution* of a variable V_i given evidence e, i.e.:

$$P(V_i \mid e)$$

for a given Bayesian network with associated probability distribution P

Problem solving using Bayesian networks:

- Classification: $P(V_i \mid e)$ with V_i class variable
- Decision making: $P(V_i \mid e, d)$ with d decision variable
- (Bayesian) learning: $\Pr(M \mid D)$, with M a BN model and D data

Notation: V_i's will denote variables and vertices (nodes) at the same time

Naive inference

- We can perform inference by using two rules:
  - conditioning

$$P(V_i \mid e) = \frac{P(V_i, e)}{P(e)}$$

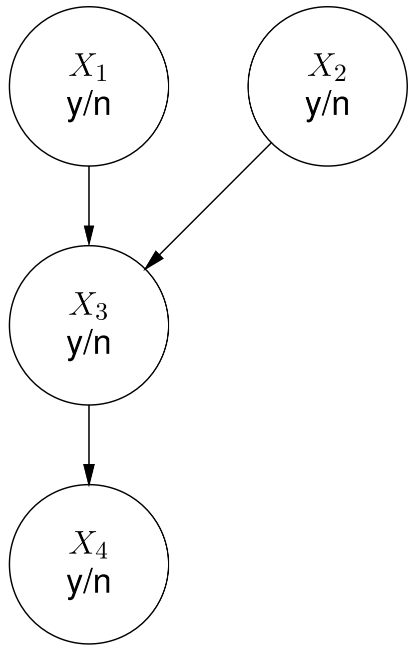
  - marginalisation

$$P(V_i) = \sum_{V(G) \setminus \{V_i\}} P(V_1, \ldots, V_n)$$

- Using factorisation of a Bayesian network

$$P(V(G)) = P(V_1, \ldots, V_n) = \prod_{i=1}^{n} P(V_i \mid \mathrm{pa}(V_i))$$

Naive probabilistic reasoning: evidence



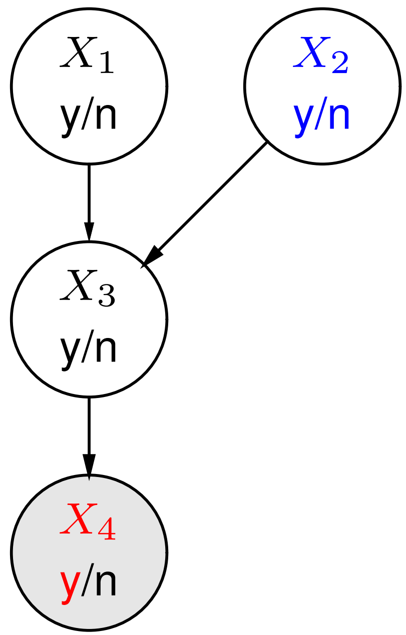
$P(x_4 \mid x_3) = 0.4$
$P(x_4 \mid \neg x_3) = 0.1$
$P(x_3 \mid x_1, x_2) = 0.3$
$P(x_3 \mid \neg x_1, x_2) = 0.5$
$P(x_3 \mid x_1, \neg x_2) = 0.7$
$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$
$P(x_1) = 0.6$
$P(x_2) = 0.2$

Using naive inference: $P(x_3 \mid x_2) =$? and $P(x_2 \mid x_4)$?

- Complexity of this algorithm is $O(n \cdot 2^n)$ with $n = |V(G)|$
- Becomes computationally feasible when we use the distributive law:

$$(ab + ac) = a(b + c)$$

Naive probabilistic reasoning: evidence

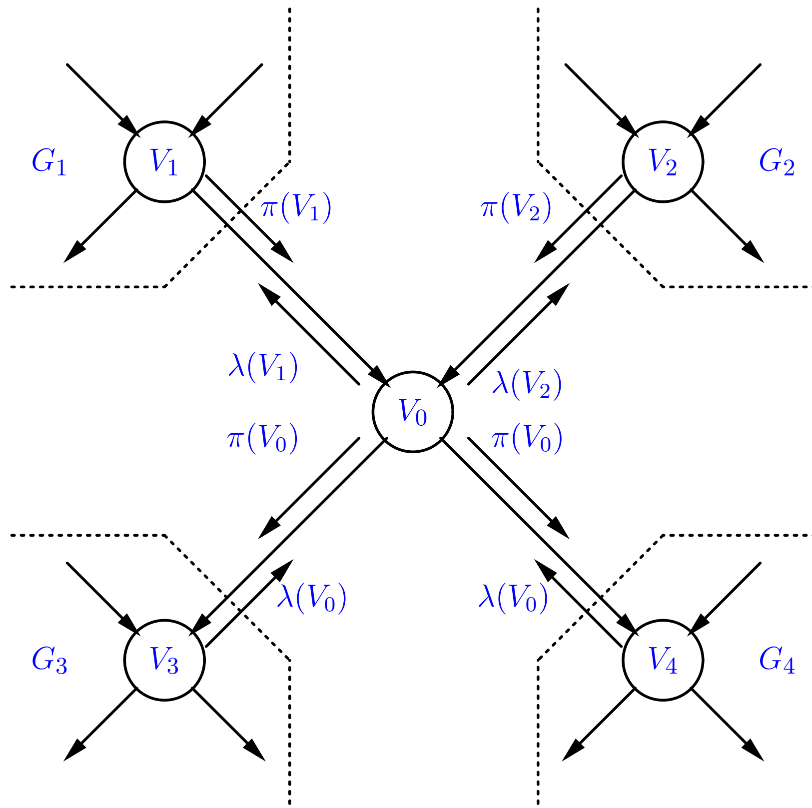


$P(x_4 \mid x_3) = 0.4$
$P(x_4 \mid \neg x_3) = 0.1$
$P(x_3 \mid x_1, x_2) = 0.3$
$P(x_3 \mid \neg x_1, x_2) = 0.5$
$P(x_3 \mid x_1, \neg x_2) = 0.7$
$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$
$P(x_1) = 0.6$
$P(x_2) = 0.2$

$$P^{\mathcal{E}}(x_2) = P(x_2 \mid x_4) = \frac{P(x_4 \mid x_2)P(x_2)}{P(x_4)} \text{ (Bayes' rule)}$$

$$= \frac{\sum_{X_3} P(x_4|X_3) \sum_{X_1} P(X_3|X_1, x_2)P(X_1)P(x_2)}{\sum_{X_3} P(x_4 \mid X_3) \sum_{X_1,X_2} P(X_3 \mid X_1, X_2)P(X_1)P(X_2)} \approx 0.14$$

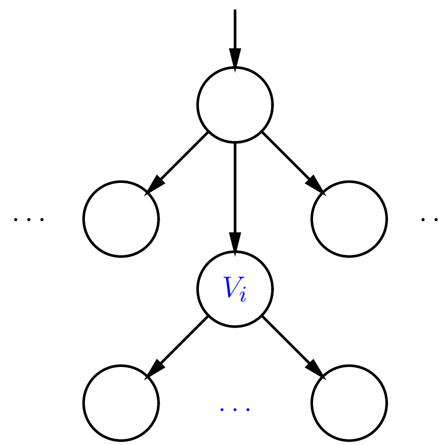
Basic idea of Pearl's algorithm



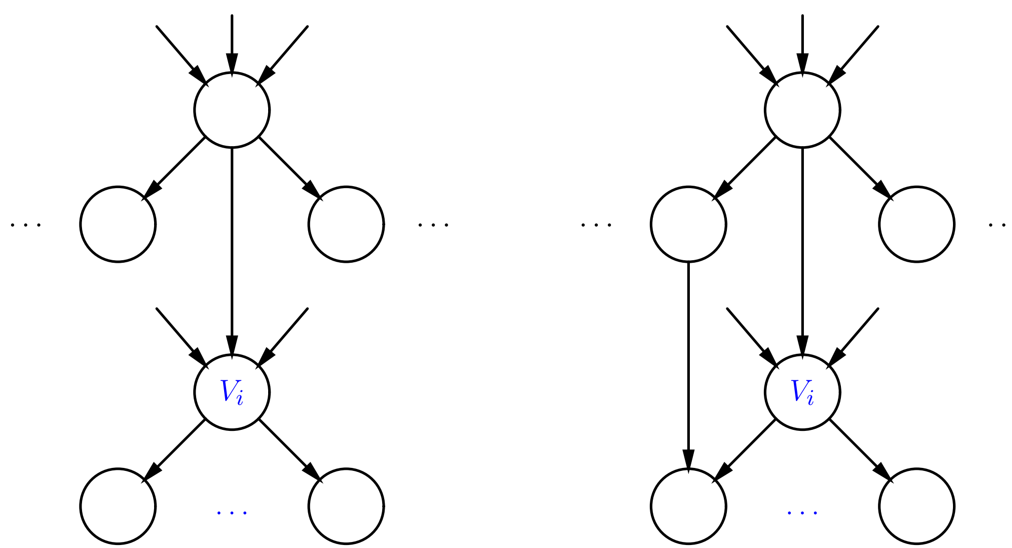
- Object-oriented approach: vertices are objects, which have local information and carry out local computations
- Updating of probability distribution by message passing: arcs are communication channels

Topology of Bayesian networks

- Directed tree:



- Singly/multiply-connected network:



singly-connected

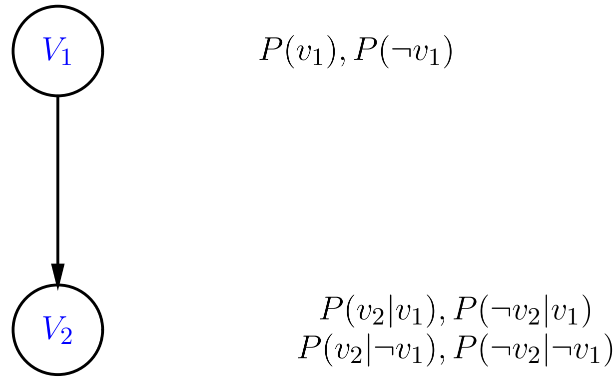
multiply-connected

Notation of messages

- Each node needs three types of parameters to compute messages and its marginal probability:
  - Causal π messages: received from parents
  - Diagnostic λ messages: received from children
  - Local memory: relevant CPT values

- π and λ messages are sent from V_i to its neighbours:
  - $\pi_{V_j}^{V_i}(V_i)$ is a message from V_i to its child V_j
  - $\lambda_{V_j}^{V_i}(V_i)$ is a message from V_j to its parent V_i

Probabilistic inference as message passing

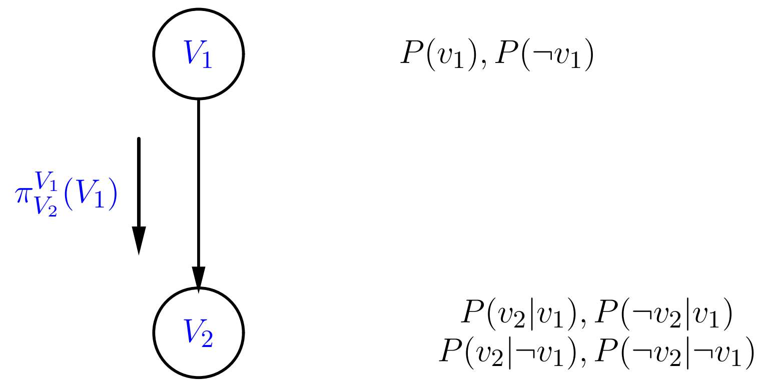


- Vertex V_1: known $P(v_1)$ and $P(\neg v_1)$
- Vertex V_2: known $P(V_2|V_1)$
- It holds that:

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

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

Message passing: causal parameter $\pi_{V_j}^{V_i}$



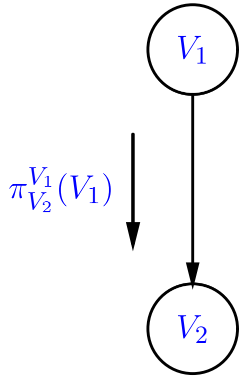
It holds that: $\pi_{V_2}^{V_1}(v_1) = P(v_1)$ and $\pi_{V_2}^{V_1}(\neg v_1) = P(\neg v_1)$

Local computation in V_2:

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

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

Example: causal parameter $\pi_{V_j}^{V_i}$



$P(v_1) = 0.8, P(\neg v_1) = 0.2$

$P(v_2|v_1) = 0.4, P(\neg v_2|v_1) = 0.6$
$P(v_2|\neg v_1) = 0.9, P(\neg v_2|\neg v_1) = 0.1$

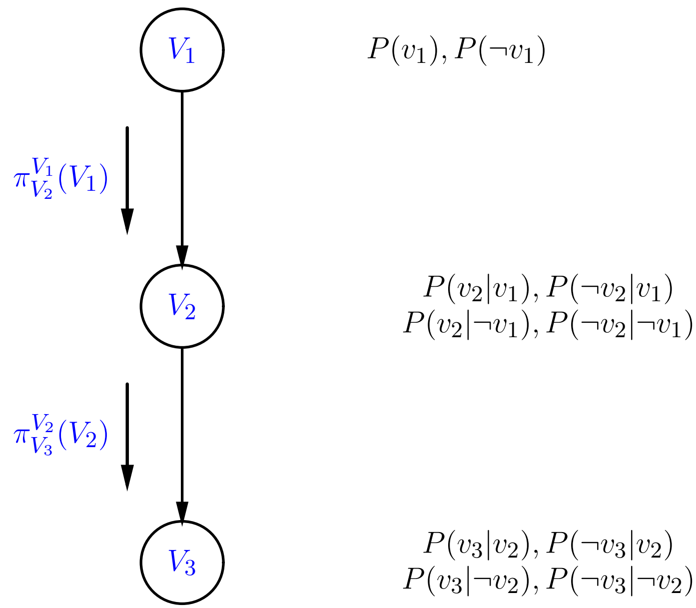
We have: $\pi_{V_2}^{V_1}(v_1) = P(v_1) = 0.8$ and $\pi_{V_2}^{V_1}(\neg v_1) = P(\neg v_1) = 0.2$

Local computation in V_2:

$$\begin{aligned} P(v_2) &= P(v_2|v_1)\pi_{V_2}^{V_1}(v_1) + P(v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1) \\ &= 0.4 \times 0.8 + 0.9 \times 0.2 = 0.5 \end{aligned}$$

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

Message passing: three vertices



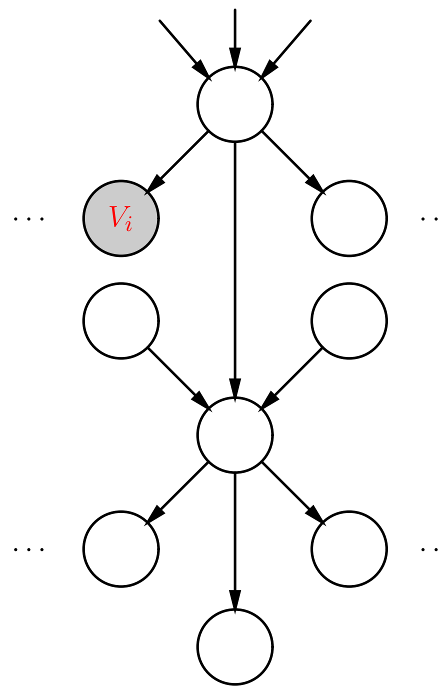
It holds that: $\pi^{V_2}_{V_3}(v_2) = P(v_2)$ and $\pi^{V_2}_{V_3}(\neg v_2) = P(\neg v_2)$

Local computation in V_3:

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

Evidence propagation

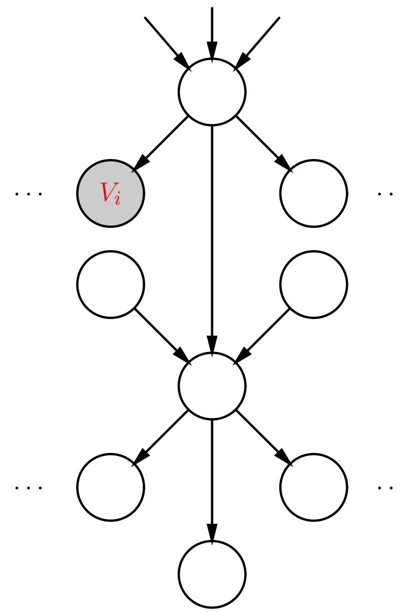
Let $\mathcal{B} = (G, P)$ be a Bayesian network with digraph G and joint probability distribution P



- Evidence is an assignment of value to a variable (i.e., instantiating the variable): $V_i =$ *true* $(= v_i)$ or $V_i =$ *false* $(= \neg v_i)$ for binary variable V_i

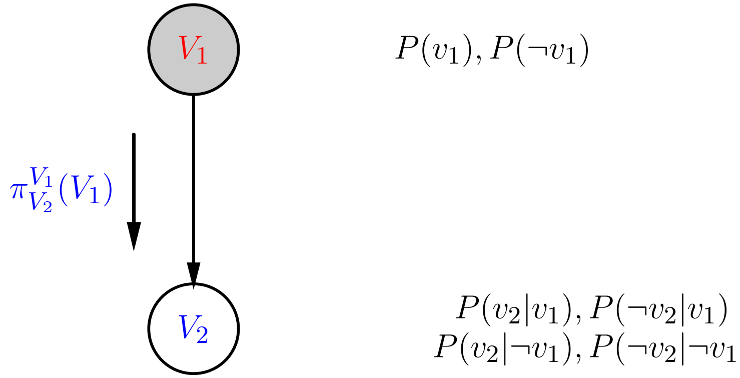
Evidence propagation (cont)

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



- Given an evidence, P no longer holds and must be updated to a new probability distribution P^*. E.g., for evidence v_i it holds that $P^*(v_i) = 1$ ($P^*(\neg v_i) = 0$), whereas originally $P(v_i) = 0.3$ ($P(\neg v_i) = 0.7$).
- Entire Bayesian network must be updated

Evidence and causal parameter



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

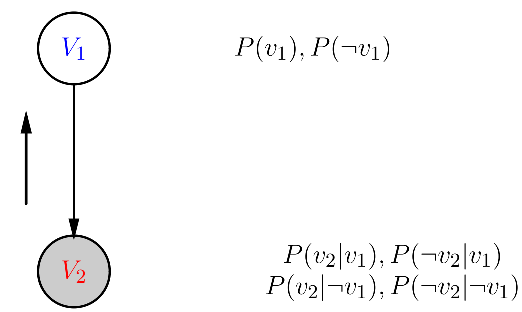
$$\pi_{V_2}^{V_1}(v_1) = 1, \ \pi_{V_2}^{V_1}(\neg v_1) = 0$$

Local computation in V_2:

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

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

Evidence and diagnostic parameter



Evidence: assume that $V_2 =$ *true* $(= v_2)$

$$P^*(v_2) = 1 \text{ , } P^*(\neg v_2) = 0$$

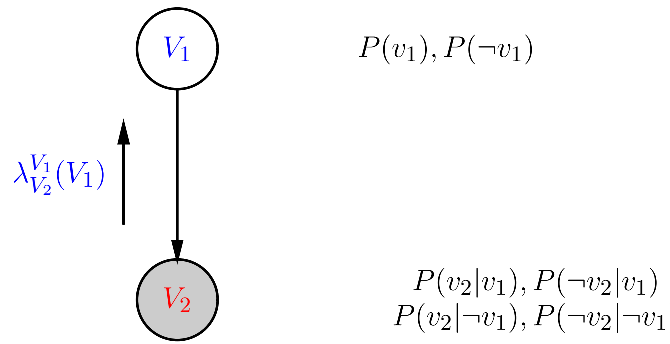
Updated probability distribution $P^*(V_1)$:

$$P^*(v_1) = P(v_1|v_2) = \frac{P(v_2|v_1)P(v_1)}{P(v_2)}$$

$$P^*(\neg v_1) = P(\neg v_1|v_2) = \frac{P(v_2|\neg v_1)P(\neg v_1)}{P(v_2)}$$

for which V_1 needs $P(V_2|V_1)$ from V_2: message $\lambda_{V_2}^{V_1}(V_1)$

Evidence and diagnostic parameter (cont)



Evidence: assume that $V_2 =$ *true* $(= v_2)$

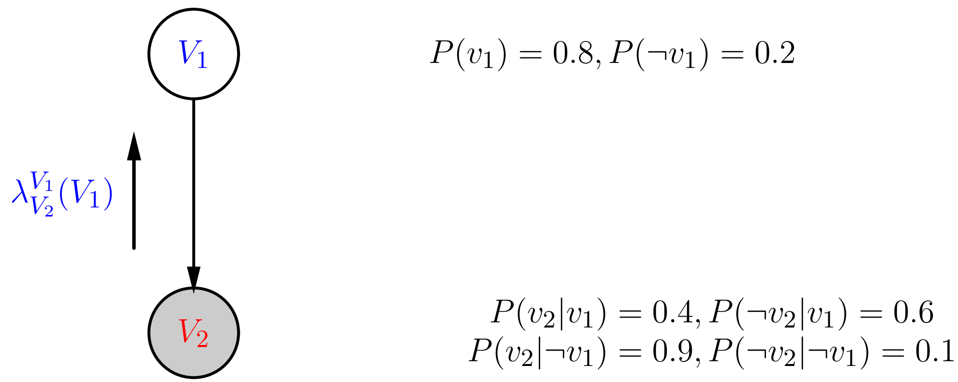
V_2 sends a message $\lambda_{V_2}^{V_1}(V_1)$ to V_1 so V_1 can compute $P^*(V_1)$

This message is defined as follows:

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

$\lambda_{V_2}^{V_1}(V_1)$ is called the diagnostic parameter

Example: diagnostic parameter



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

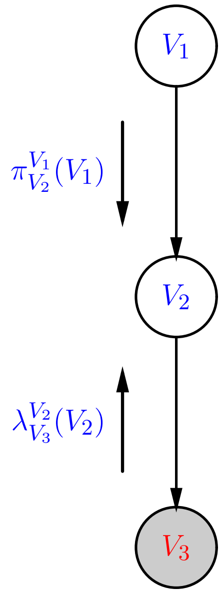
$$P^*(v_1) = \frac{P(v_2|v_1)P(v_1)}{P(v_2)} = \alpha\, \lambda_{V_2}^{V_1}(v_1)P(v_1)$$

$$= \alpha \times 0.4 \times 0.8 = 0.32\alpha$$

$$P^*(\neg v_1) = \frac{P(v_2|\neg v_1)P(\neg v_1)}{P(v_2)} = \alpha\, \lambda_{V_2}^{V_1}(\neg v_1)P(\neg v_1)$$

$$= \alpha \times 0.9 \times 0.2 = 0.18\alpha$$

Causal and diagnostic parameters combined



$P(v_1) = 0.8, P(\neg v_1) = 0.2$

$P(v_2|v_1) = 0.4, P(\neg v_2|v_1) = 0.6$
$P(v_2|\neg v_1) = 0.9, P(\neg v_2|\neg v_1) = 0.1$

$P(v_3|v_2) = 0.7, P(\neg v_3|v_2) = 0.3$
$P(v_3|\neg v_2) = 0.5, P(\neg v_3|\neg v_2) = 0.5$

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

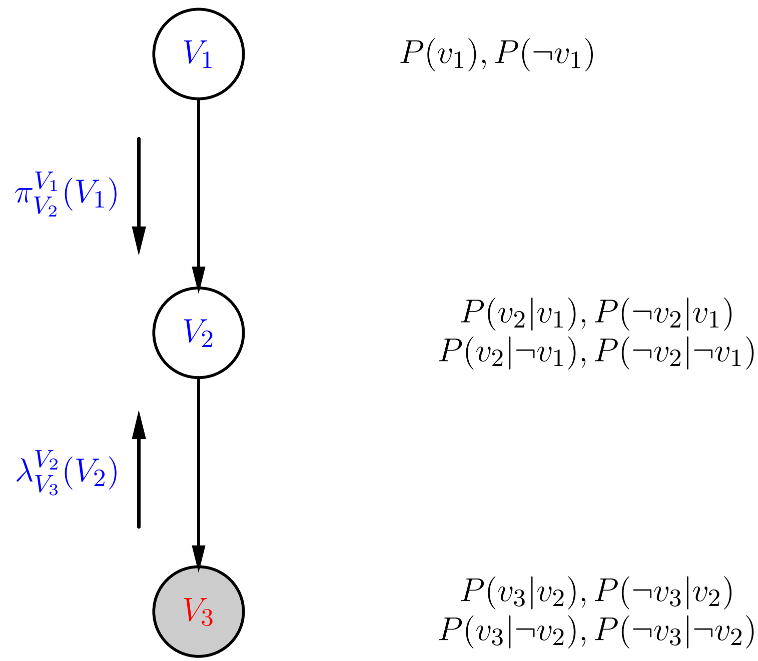
$$P^*(v_2) = \alpha\, \lambda_{V_3}^{V_2}(v_2)[P(v_2|v_1)\pi_{V_2}^{V_1}(v_1) + P(v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1)]$$

$$= \alpha \times 0.7[0.4 \times 0.8 + 0.9 \times 0.2] = 0.35\alpha$$

$P^*(\neg v_2)$ = analogous = 0.25α (thus, $\alpha = 1\frac{2}{3}$)

$\lambda_{V_3}^{V_2}(V_2) = P(V_3|V_2)$, and $\pi_{V_2}^{V_1}(V_1) = P(V_1)$

Towards a generic formula



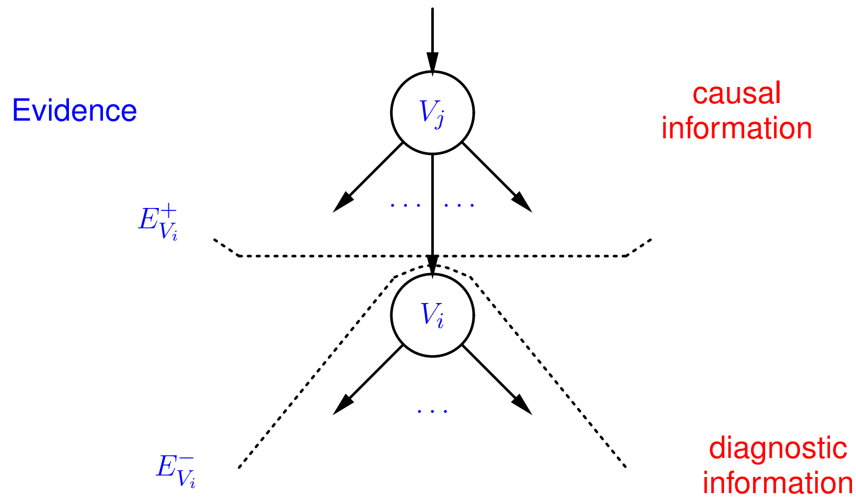
Updated probability distribution $P^*(V_2)$ for *evidence* v_3:

$$P^*(V_2) = \alpha\, \lambda^{V_2}_{V_3}(V_2)[P(V_2|v_1)\pi^{V_1}_{V_2}(v_1) + P(V_2|\neg v_1)\pi^{V_1}_{V_2}(\neg v_1)]$$

= $\alpha \cdot$ diagnostic information for $V_2 \cdot$ causal information for V_2

= $P(V_2 \mid \text{Evidence})$

Generic formula: data fusion



Data fusion lemma:

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

where:

- e: evidence
- α: normalisation constant
- $\pi(V_i) \triangleq P(V_i \mid e^+_{V_i})$: *compound* causal parameter
- $\lambda(V_i) \triangleq P(e^-_{V_i} \mid V_i)$: *compound* diagnostic parameter

Compound parameters: basic questions

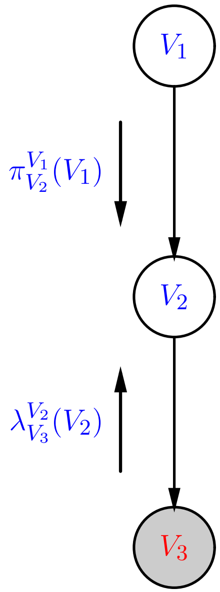
- What is the compound causal parameter $\pi(v_i)$ of a node V_i if $e^+_{V_i} = \varnothing$?

- What is the compound diagnostic parameter $\lambda(v_i)$ of a node V_i if $e^-_{V_i} = \varnothing$? And $\lambda(\neg v_i)$?

- If the evidence e consists of $\{v_i\}$, what is the value of the normalisation constant α in the data fusion lemma:

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

Example of data fusion



$P(v_1) = 0.8, P(\neg v_1) = 0.2$

$P(v_2|v_1) = 0.4, P(\neg v_2|v_1) = 0.6$
$P(v_2|\neg v_1) = 0.9, P(\neg v_2|\neg v_1) = 0.1$

$P(v_3|v_2) = 0.7, P(\neg v_3|v_2) = 0.3$
$P(v_3|\neg v_2) = 0.5, P(\neg v_3|\neg v_2) = 0.5$

Evidence v_3:

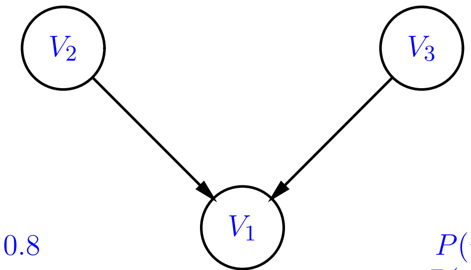
- $\lambda(v_2) = \lambda_{V_3}^{V_2}(v_2) = 0.7$
- $\pi(v_2) = P(v_2|v_1)\pi_{V_2}^{V_1}(v_1) + P(v_2|\neg v_1)\pi_{V_2}^{V_1}(\neg v_1) = 0.5$
- $\alpha = 1\frac{2}{3}$

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

Simple network example

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

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



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

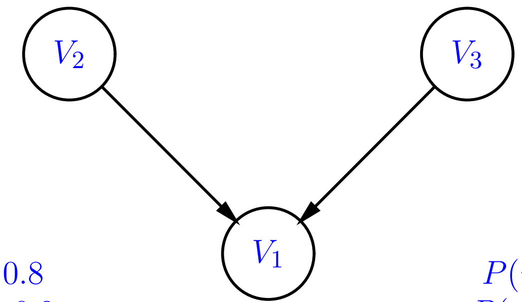
$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

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

Simple network example (cont)

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

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



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

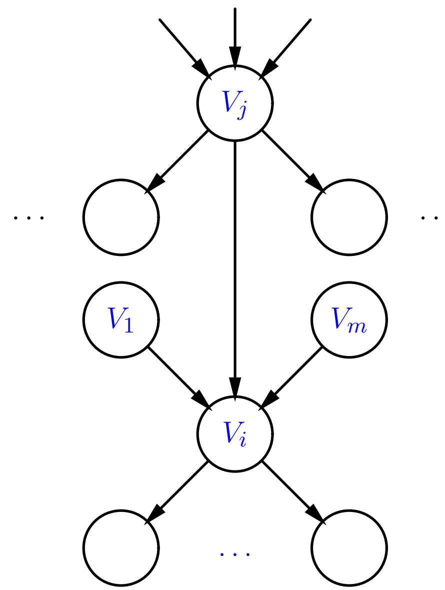
$P(\neg v_1|v_2 \wedge v_3) = 0.2$
$P(\neg v_1|\neg v_2 \wedge v_3) = 0.1$
$P(\neg v_1|v_2 \wedge \neg v_3) = 0.5$
$P(\neg v_1|\neg v_2 \wedge \neg v_3) = 0.4$

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

Given that no evidence is provided, $\lambda(v_1) = 1$.
Analogously, $\lambda(\neg v_1) = 1$.
Therefore,

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

Compound causal parameter for SCN

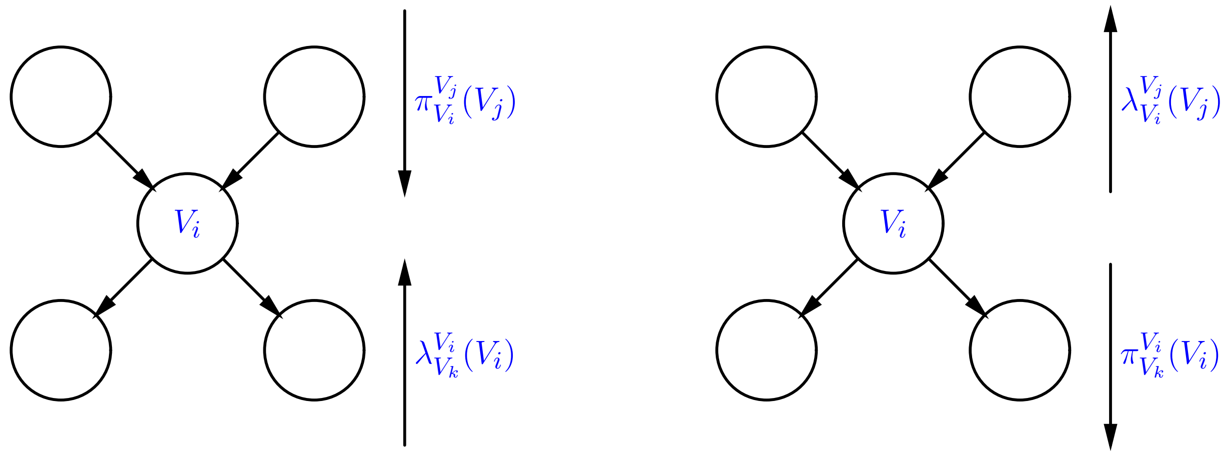


Define $\pi_{V_i}^{V_j}(V_j) \triangleq P(V_j \mid e_{V_j}^{+}, e_{\text{children}(V_j)\neq V_i}^{-}, e_{V_j})$. It follows:

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

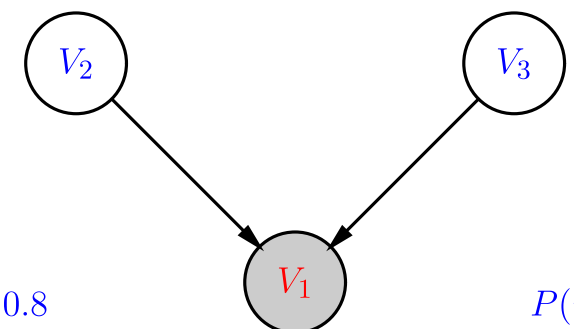
with parents $\text{pa}(V_i) = V_1 \wedge \cdots \wedge V_j \wedge \cdots \wedge V_m$

Messages to children and parents



Example:

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



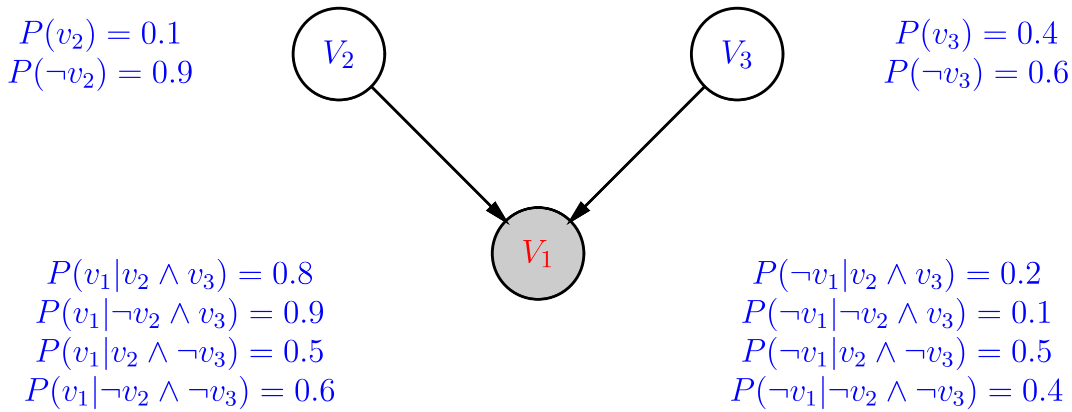
$P(v_3) = 0.4$
$P(\neg v_3) = 0.6$

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

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

- Suppose that the evidence $V_1 = true$ is observed, we want to compute the updated probability of V_2

Messages to child and parent (cont)



The probabilities of interest are computed according to the fusion lemma:

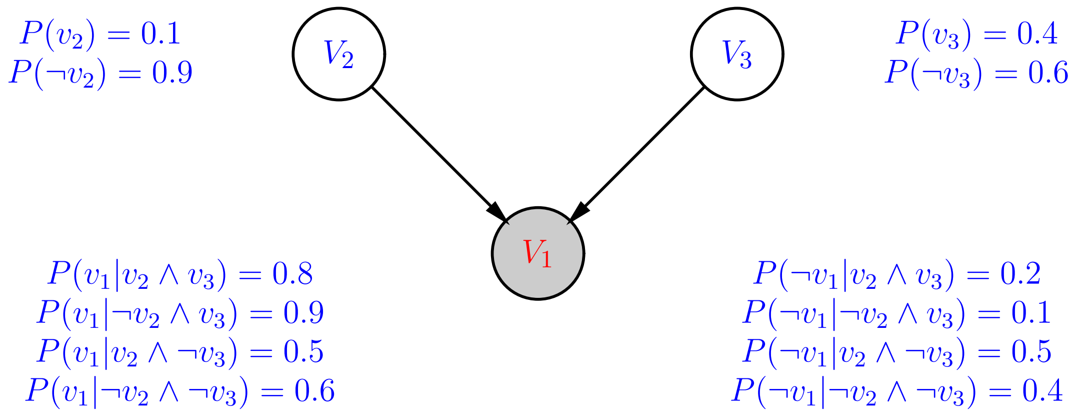
$$P^*(V_2) = \alpha \cdot \pi(V_2) \cdot \lambda(V_2)$$

V_2 has now to compute its compound parameters

Having no parents, the compound causal parameter for V_2 is then:

$$\pi(v_2) = P(v_2)$$
$$\pi(\neg v_2) = P(\neg v_2)$$

Messages to child and parent (cont)



The values of the compound diagnostic parameter are calculated from

$$\lambda(v_2) = \lambda_{V_1}^{V_2}(v_2)$$

$$\lambda(\neg v_2) = \lambda_{V_1}^{V_2}(\neg v_2)$$

From its sucessor V_1, vertex V_2 receives the diagnostic parameter $\lambda_{V_1}^{V_2}$ which should thus be equal to $P(V_2 \mid v_1)$

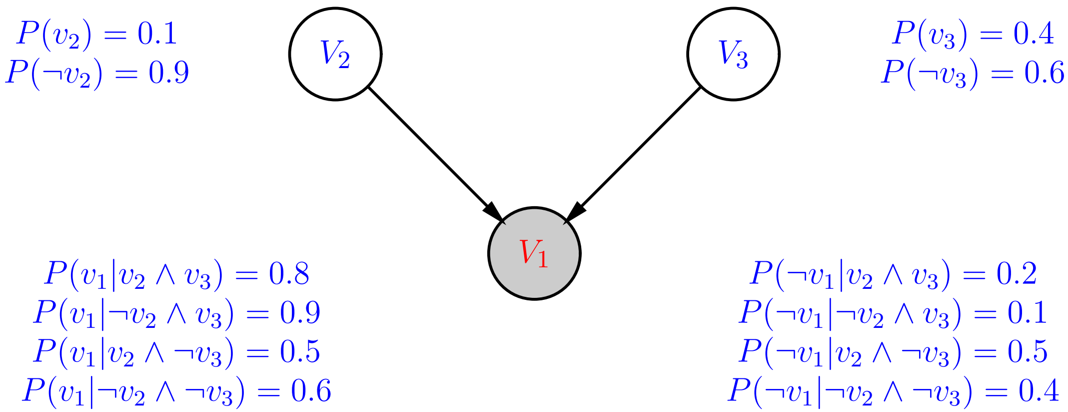
Messages to child and parent (cont)

Such diagnostic parameter has values:

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

Analogously for $\lambda_{V_1}^{V_2}(\neg v_2) = 0.72$

Messages to child and parent (cont)



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

$\lambda_{V_1}^{V_2}(v_2) = 0.62$ and $\lambda_{V_1}^{V_2}(\neg v_2) = 0.72$

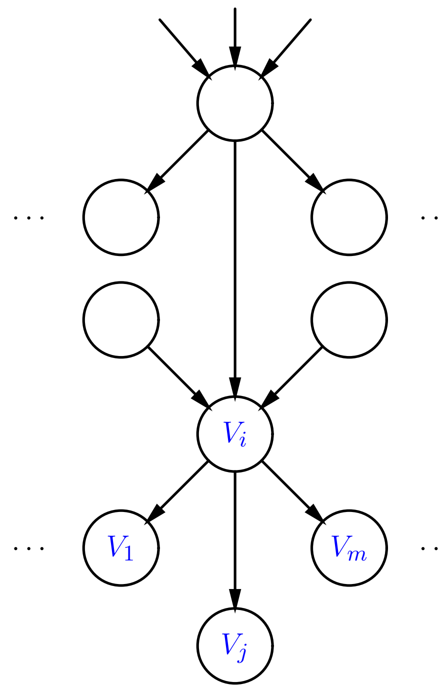
$\pi(v_2) = 0.1$ and $\pi(\neg v_2) = 0.9$

Result:

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

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

Compound diagnostic parameter for SCN



Define $\lambda^{V_i}_{V_j}(V_i) \triangleq P(e^-_{V_j}, e^+_{\mathrm{pa}(V_j)\neq V_i} \mid V_i)$. If V_i is not observed:

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

Summary of local computations

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

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

$$\lambda(V_i) = \prod_{j=1}^{m} \lambda_{V_j}^{V_i}(V_i) \quad \text{if } V_i \notin E$$

$$\pi_{V_j}^{V_i}(V_i) = \alpha \cdot \pi(V_i) \cdot \prod_{k \neq j} \lambda_{V_k}^{V_i}(V_i) \quad \text{if } V_i \notin E$$

$$\lambda_{V_j}^{V_i}(V_i) = \beta \sum_{V_j} \lambda(V_j) \cdot \sum_{V_k \in \mathrm{pa}(V_j), k \neq i} P(V_j \mid \mathrm{pa}(V_j)) \prod_{k \neq i} \pi_{V_j}^{V_k}(V_k)$$

Algorithm steps

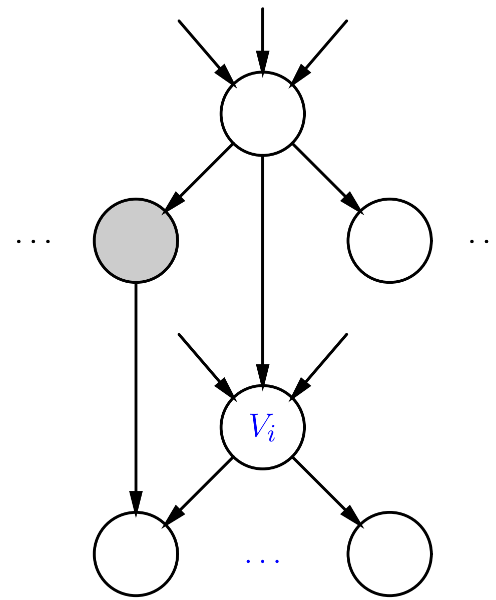
In each iteration, each node V_i does the following:

- if V_i has received all the causal messages from its parents, compute $\pi(V_i)$
- if V_i has received all the λ messages from its children, compute $\lambda(V_i)$
- if $\pi(V_i)$ is known, and V_i received all the λ messages from its children except for V_j, compute $\pi_{V_j}^{V_i}(V_i)$ and send it to V_j
- if $\lambda(V_i)$ is known and V_i received all the π messages from all parents except for V_j, compute $\lambda_{V_i}^{V_j}(V_j)$ and send it to V_j

Overview of Pearl's algorithm

- All the computations are local
- Efficient for local computation property and parallel, distributed implementations
- However, there is a summation over all joint instantiations of parent nodes $\Rightarrow$ exponential in the number of parents
  - if parents sets are bound in size by a constant, the runtime is *linear*
- Therefore, computationally infeasible in networks where nodes have too many parents
- Number of data propagation cycles proportional to the length of path(s) from evidence node(s)

Multiply-connected networks inference



- At least two nodes are connected by more than one path (in the underlying undirected path)
- Thus, some variables can influence another through more than one causal mechanism
- And same evidence counted more than once

$\Rightarrow$ next week more

References

- Kim, J. H. and Pearl, J. (1983) A computational model for causal and diagnostic reasoning in inference systems. In *Proceeding of the Eighth International Joint Conference in Artificial Intelligence (IJCAI)*, pp. 190-193.
- Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*. Morgan Kauffman. ISBN 0-934613-73-7.
- Spiegelhalter, D. J. and Lauritzen, S. L. (1990) Sequential updating of conditional probabilities on directed graphical structures. *Networks*, **20**, pp. 579-605.
- Jensen, F. V. and Nielsen, T. D. *Bayesian Networks and Decision Graphs*. Springer. ISBN 978-0-387-68281