

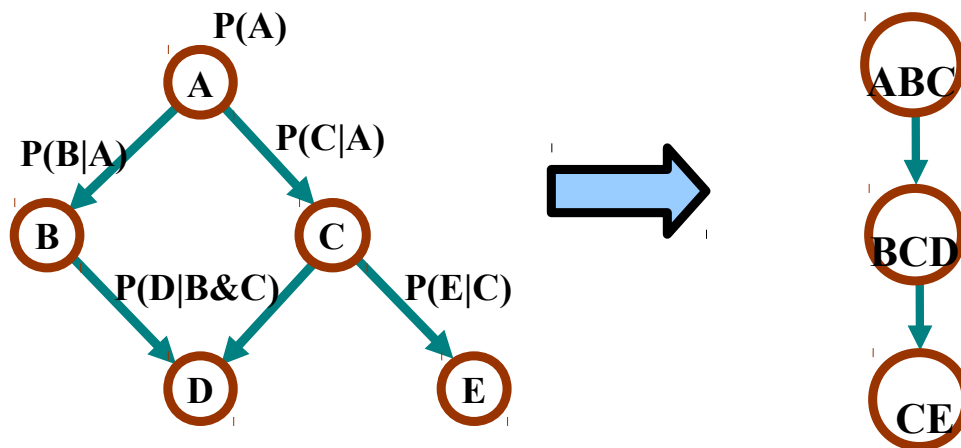
# Lecture 10:

## Propagating Probabilities in a Join Tree

# The story so far

Data sets with high inter-variable dependency produce Bayesian networks with lots of loops. This is bad news for probability propagation.

For these cases we aim to transform the network into a singly connected “join tree” where probability propagation is possible.



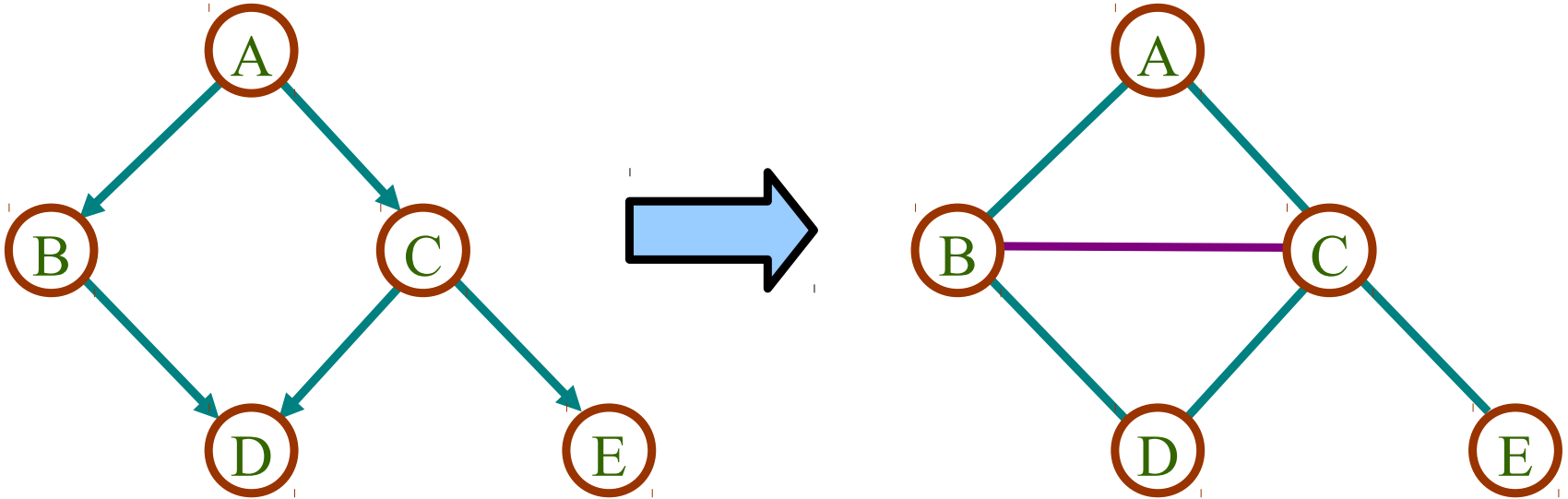
# Building a join tree

A join tree can always be found from a given causal graph as follows:

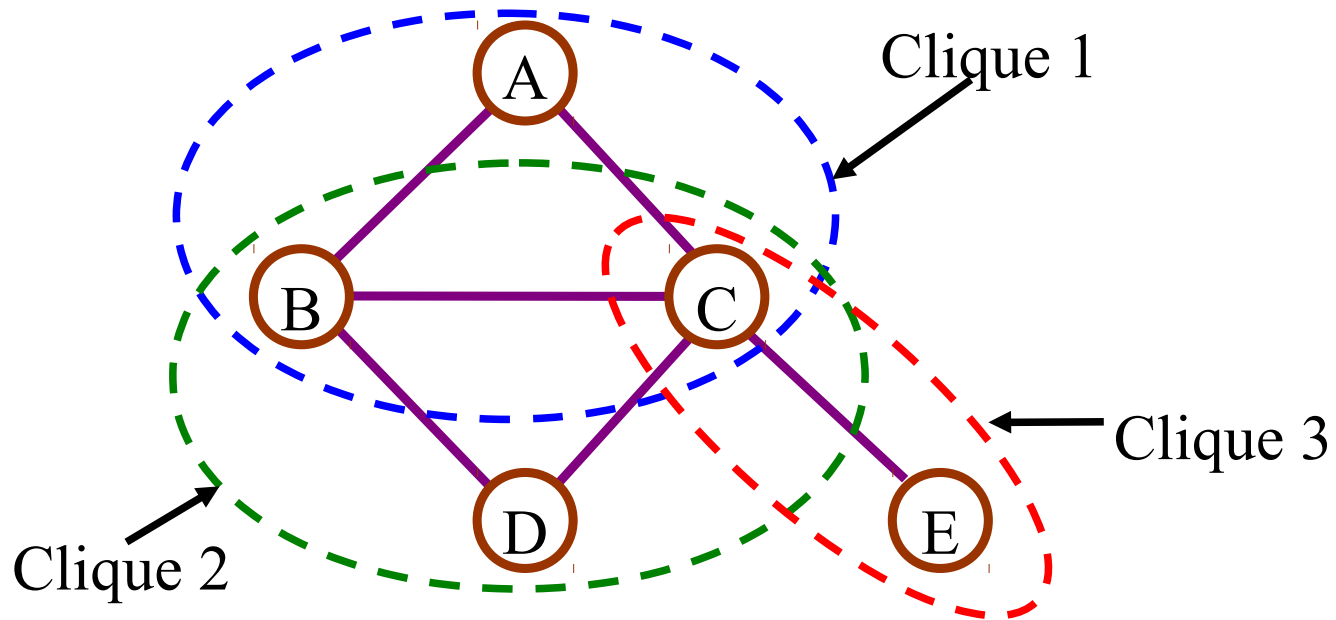
1. Moralise the graph (join any unjoined parents)
2. Triangulate the graph
3. Identify the cliques of the resulting graph
4. For each clique  $C_i$  find the set of its variables  $\{X_i\}$  whose parents  $Pa(X_i)$  also belong to the clique. This is normally the R set for the clique.
5. Initialise the clique potential function  $\psi$  as follows:

$$\psi(C_i) = \psi(W_i) = \prod_{\{X_i\}} P(X|Pa(X))$$

# Moralising and triangulation



# Finding the Cliques



# Dividing the variables in Cliques

Each original variable must be in the  $R$  set of exactly one clique and its original parents must be in the same clique. The moralisation operation ensures that this is the case.

We use these constraints to determine the  $R$  and  $S$  sets for each clique.

# Initialising Potential functions

The potential functions can then be found for each clique. They are the product of the conditional or prior probability tables with distributions over the variables in the R set:

Index	W	S	R	$\Psi$
1	$ABC$		$ABC$	$P(A)P(B A)P(C A)$
2	$BCD$	$BC$	$D$	$P(D B\&C)$
3	$CE$	$C$	$E$	$P(E C)$

# Defining the tree of cliques

We now find an ordering of the cliques with the running intersection property.

For each clique we search through all the other cliques to find any that contain all its  $S$  set variables. Any of these can be its parent.

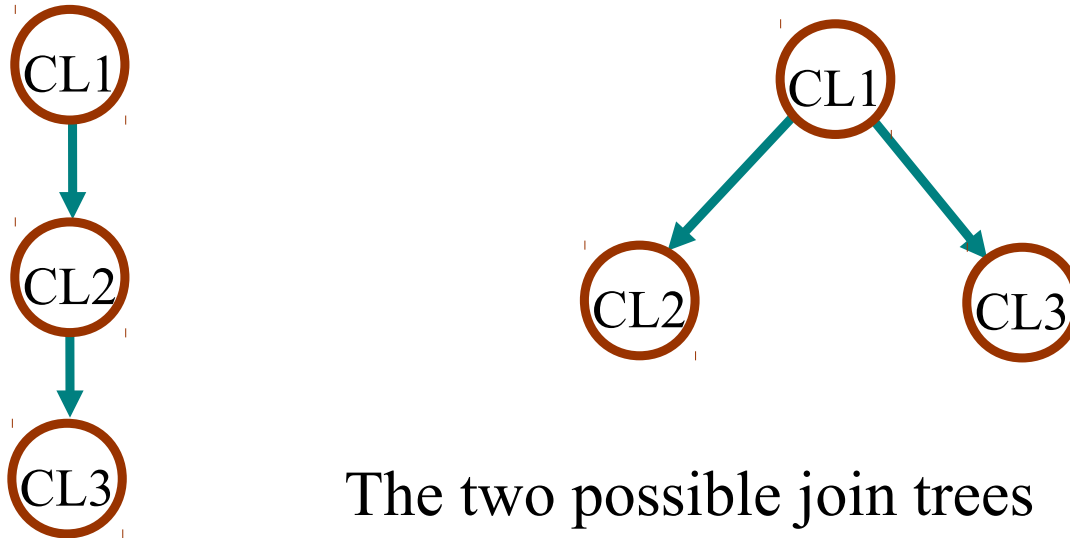
One way to proceed is to start from a clique that contains a root node and link downwards.



# The Join tree

The join tree is now completed.

In all future probability calculations we use just the join tree, disregarding our original network



The two possible join trees

# States of a clique

Each node (or clique) in a join tree is made up of several variables, and therefore will have a state for each possible combination of those variables.

Given that clique 3 is made up of two variables C and E

C has states c1 and c2

E has states e1 and e2

then:

Clique 3 has states  $\{c1,e1\}$ ,  $\{c1,e2\}$ ,  $\{c2,e1\}$ ,  $\{c2,e2\}$

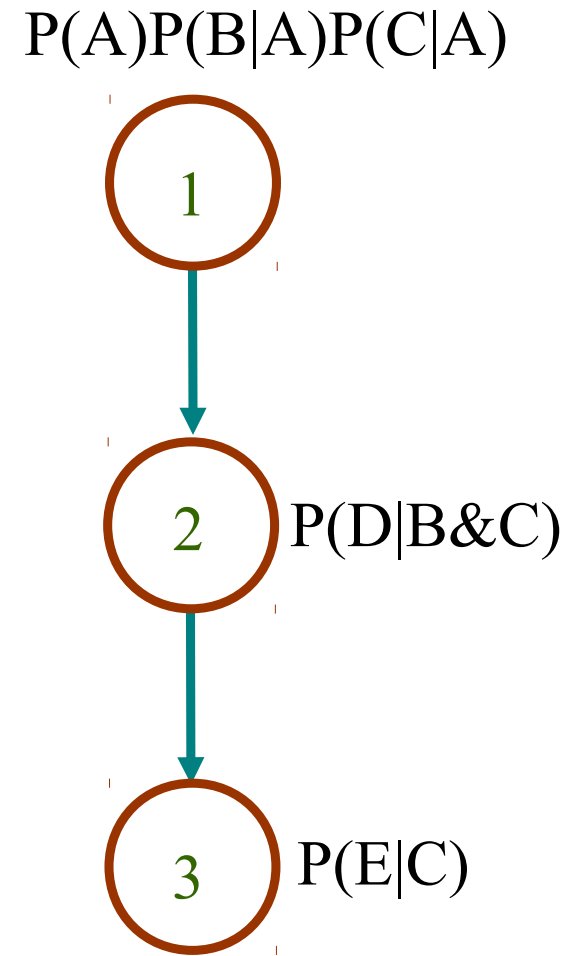
# Summary of the join tree

Clique	R	S	States
1	A,B,C	none	$\{a1,b1,c1\}$ $\{a1,b1,c2\}$ $\{a1,b2,c1\}$ $\{a1,b2,c2\}$ $\{a2,b1,c1\}$ $\{a2,b1,c2\}$ $\{a2,b2,c1\}$ $\{a2,b2,c2\}$
2	D	B,C	$\{b1,c1,d1\}$ $\{b1,c2,d1\}$ $\{b2,c1,d1\}$ $\{b2,c2,d1\}$ $\{b1,c1,d2\}$ $\{b1,c2,d2\}$ $\{b2,c1,d2\}$ $\{b2,c2,d2\}$
3	E	C	$\{c1,e1\}$ $\{c1,e2\}$ $\{c2,e1\}$ $\{c2,e2\}$

# Join Tree Probability Tables

Each node in the join tree has a “potential” function  $\psi(W)$  which allocates a probability to each state of the node.

This is initialised using the link matrices of the original Bayesian network.



# Calculating initial potential tables

Given that clique 1 has the potential function:

$$\psi(C11) = \psi(W1) = P(A)P(B|A)P(C|A)$$

we calculate a potential table with an entry for each state, thus:

$$\psi(a1,b1,c1) = P(a1)P(b1|a1)P(c1|a1)$$

$$\psi(a1,b1,c2) = P(a1)P(b1|a1)P(c1|a2)$$

etc.

# Calculating the probabilities

The basis of the propagation algorithm is to find the joint probabilities of the cliques for a given instantiation.

Initially we set the potential functions to:

$$\psi(C11) = \psi(W1) = P(A)P(B|A)P(C|A)$$

$$\psi(C12) = \psi(W2) = P(D|B\&C)$$

$$\psi(C13) = \psi(W3) = P(E|C)$$

And we compute

$$P(C11) = P(W1) = P(A\&B\&C)$$

$$P(C12) = P(W2) = P(B\&C\&D)$$

$$P(C13) = P(W3) = P(C\&E)$$

# Computing probabilities of the cliques

Each clique has a set of variables, and we have divided them into those that have been seen higher in the tree (S) and those that have not (R)

Clique	Variables (W)	Seen (S)	New (R)
1	A,B,C	None	A,B,C
2	B,C,D	B,C	D
3	C,E	C	E

The probability of a clique can be written:

$$P(Cli) = P(Wi) = P(Ri \& Si) = P(Ri | Si) P(Si)$$

# Overview

The algorithm consists of two steps:

1. Start at the bottom of the tree and propagate  $\lambda$  messages up to the top. And at each stage we calculate  $P(R|S)$ .
2. Go from the top of the tree to the bottom propagating  $\pi$  messages. At each stage we calculate  $P(S)$  and then  $P(W_i) = P(S) P(R|S)$



# Getting the variable probabilities

The original variable probabilities can be obtained by marginalisation from the clique probabilities. For example:

$$P(B) = \sum_{AC} P(A \& B \& C) \quad (\text{Summing over } A \text{ and } C)$$

for the individual states of B we have:

$$P(b_1) = P(a_1 \& b_1 \& c_1) + P(a_1 \& b_1 \& c_2) + P(a_2 \& b_1 \& c_1) \\ + P(a_2 \& b_1 \& c_2)$$

$$P(b_2) = P(a_1 \& b_2 \& c_1) + P(a_1 \& b_2 \& c_2) + P(a_2 \& b_2 \& c_1) \\ + P(a_2 \& b_2 \& c_2)$$

# Computing the P(R|S) values

For a given join tree the algorithm starts at the base of the tree and works upwards.

For our example we have initially

$$\psi(W1) = P(A)P(B|A)P(C|A)$$

$$\psi(W2) = P(D|B\&C)$$

$$\psi(W3) = P(E|C)$$

And the potential functions are always a factorisation of the joint probability of the original variables ie:

$$P(V) = \prod_i \Psi(W_i)$$



# Computing $P(R|S)$ : 1

We can calculate the value of  $P(R|S)$  from the potential function using the formula:

$$P(R|S) = \Psi(W_i) / \sum_R \Psi(W_i)$$

(The formal proof is in Neapolitan)

## Computing $P(R|S)$ : 2

In general the the clique potential function may just represent all the accumulated evidence for that node. This will be the case after instantiation of one of the clique variables.

In this case we can think of the operation:

$$\Psi(W_i) / \sum_R \Psi(W_i)$$

as a normalisation that creates the link matrix  $P(R|S)$  from the evidence accumulated so far.

## Computing $P(R|S)$ : 3

If the clique potential function is the joint probability of the clique variables ( $W_i$ ) we have that:

$$P(R|S) = P(R\&S)/P(S) = \Psi(W_i)/\sum_R \Psi(W_i)$$

as required.

If the clique potential function is already in the form  $P(R|S)$  then

$$\sum_R \Psi(W_i) = \sum_R P(R|S) = \{1, 1, \dots, 1\}$$

(Since the columns of a link matrix sum to 1)

## Sending a $\lambda$ message

The  $\lambda$  message that a clique sends to its parents is:

$$\sum_{\mathbf{R}} \Psi(\mathbf{W}_i)$$

Note that this is evidence for the  $\mathbf{S}$  variables of any clique.

The potential function of the parent is updated by multiplying it by this evidence.

# Updating the potential function of a clique

The  $\lambda$  message updates the potential function of the parent clique by multiplication. For example if it contained evidence about a variable  $C$  and

$$\lambda(C) = \{ \lambda(c1), \lambda(c2) \}$$

The entries in  $\psi(W_i)$  are multiplied by the evidence where the states correspond, thus:

$\psi(b1, d2, c1)$  is multiplied by  $\lambda(c1)$

# Maintaining the potential representation

During message passing at each node we:

1. Compute the evidence for the S variables  $\sum_R \Psi(W_i)$
2. Divide the potential function of clique by this evidence so it becomes  $P(R|S)$
3. Multiply the potential function of the parent clique by the evidence.

Note that the potential functions still are a representation of the joint probability:

$$P(V) = \prod_i \Psi(W_i)$$



# Propagation during Initialisation

Let us consider the case where none of the variables have been instantiated.

For this case we expect that all  $\lambda$  messages will contain no evidence, but there will be  $\pi$  evidence from the root nodes.

## Applying the formulae to clique 3

For the bottom clique the  $\lambda$  message is:

$$\Sigma_R \Psi(W_i) = \Sigma_E P(E|C) = \{1,1\}$$

Which is what we expect since  $P(R|S)=P(E|C)$

So the potential function of this clique and its parent remain unchanged.

## Calculating Clique 2 $P(R_2|S_2)$

As no evidence was propagated the potential function of clique 2 is unchanged and we have that:

$$\begin{aligned} P(R_2|S_2) &= \psi(W_2) / \sum_D \psi(W_2) \\ &= P(D|B\&C) / \sum_D P(D|B\&C) \\ &= P(D|B\&C) \end{aligned}$$

Again the  $\lambda$  message contains no evidence:

$$\lambda(B\&C) = \sum_D \psi(W_2) = \sum_D P(D|B\&C) = \{1,1,1,1\}$$

# The Root clique

Clique 1 has  $R = \{A,B,C\}$  and  $S = \{\}$

Initially  $\psi(W1) = P(A,B,C) = P(A)P(B|A)P(C|A)$

Normally the potential function would be multiplied by the messages from below. In this case nothing changes as the message contained no evidence.

Again it turns out that  $P(R|S) = \psi(W1)$ .

# Finding the $P(S_i)$ values

The  $P(S_i)$  values are found from the top downwards.

(In effect they are  $\pi$  messages)

Clique 1 has  $S_1 = \{\}$ , so  $P(S_1) = 1$

Clique 2 has  $S_2 = \{B \& C\}$  but the probability of clique 1 is  $P(A \& B \& C)$ . Thus we can calculate  $P(S_2)$  by marginalisation:

$$P(S_2) = P(B \& C) = \sum_A P(A \& B \& C)$$

eg

$$P(b_1 \& c_2) = P(a_1 \& b_1 \& c_2) + P(a_2 \& b_1 \& c_2) \text{ etc}$$

# Downward propagation

The value of  $P(S_2)$  (which contains evidence) allows us to compute the probability of clique 2:

$$P(C_{12}) = P(R_2|S_2) P(S_2) = P(B\&C\&D)$$

we now marginalise to find  $P(S_3)$

$$P(S_3) = P(C) = \sum_{BD} P(B\&C\&D)$$

Now all the clique probabilities have been found

# Instantiation

If we instantiate a node in a causal network this will change the potential tables in any node of the tree containing the instantiated variable.

Let us consider instantiating the node D in our example. Suppose our initial potential table was:

	b1&c1	b1&c2	b2&c1	b2&c2
d1	0.8	0.9	0.7	0.05
d2	0.2	0.1	0.3	0.95

## The new potential function of clique 2

Since we know  $D=d1$  all the entries for  $D=d2$  become zero. Just the values in the old potential function for which  $D=d1$  remain, for example:

	b1&c1	b1&c2	b2&c1	b2&c2
d1	0.8	0.9	0.7	0.05
d2	0	0	0	0



## Calculating the values of $P(R3|S3)$

Having instantiated  $D$  we proceed as before, and again we find that clique 3 is unchanged so:

$$\psi(W3) = P(E|C) = P(R3|S3)$$

As before the message to be propagated upwards will be  $\{1,1\}$ , no evidence.

We proceed to clique 2 without changing anything

## Calculating the $\lambda$ message at clique 2

Using  $\lambda = \Sigma_D \psi(W2)$  we now have:

	b1&c1	b1&c2	b2&c1	b2&c2
d1	0.8	0.9	0.7	0.05
d2	0	0	0	0
$\Sigma_D P(D B\&C)$	0.8	0.9	0.7	0.05

Thus we have the  $\lambda$  message

$$\lambda(B\&C) = \{0.8, 0.9, 0.7, 0.05\}$$

This evidence is from the instantiation of D which is not seen elsewhere in the tree

## Calculating $P(R_2|S_2)$

Having computed the  $\lambda$  message we can now update the potential function of clique 2 by dividing out the  $\lambda$  evidence to leave  $P(R|S)$

$$P(R_2|S_2) = \psi(W_2) / \sum_{R_2} \psi(W_2) = \psi(W_2) / \psi(W_2)$$

$$=$$

	$b_1 \& c_1$	$b_1 \& c_2$	$b_2 \& c_1$	$b_2 \& c_2$
$d_1$	1	1	1	1
$d_2$	0	0	0	0

This replaces the old potential function

## At the root node:

We now need to combine the evidence to find the posterior probability of the root clique, given:

$\psi(W1) = P(A)P(B|A)P(C|A)$  from the original link matrices and

$\lambda(B\&C)$  evidence over the states of B&C following the instantiation of D

$$\psi'(W1) = \psi(W1) * \lambda(B\&C)$$

# Getting the probability of Clique 1

We now just need to normalise the clique 1 potential function to give us the joint probability of clique 1:

$$P'(W1) = \psi'(W1) / \sum_A \sum_B \sum_C \psi'(W1)$$

And now the rest of the propagation is exactly as before:

$$P'(S2) = P'(B\&C) = \sum_A P'(A\&B\&C) = \sum_A P'(W1)$$

## Getting the probability of Clique 2:

$$P'(W2) = P'(R2|S2) P'(S2)$$

and since, as we saw previously,

$$P'(R2|S2) = 1 \text{ for } D=d1$$

and 0 for  $D=d2$

so  $P'(W2)$  contains just the entries from  $P'(S2)$

we can now calculate

$$P'(S3) = P'(C) = \sum_{BD} P'(W2) (= \sum_{AB} P'(W1))$$

## Getting the probability of clique 3

We previously calculated  $P(R3|S3)$  which is just:

$P(E|C)$  (the link matrix from the original tree)

So  $P'(W3) = P'(R3|S3) * P'(S3) = P(E|C) * \sum_B P'(W2)$

# The probabilities of the variables

As previously noted we can compute the posterior probabilities of the variables by marginalisation:

For example:

$$P'(C) = \sum_E P'(W3)$$

or

$$P'(C) = \sum_{AB} P'(W1)$$



## Further instantiation

Having completed the traversal of the tree we have that the potential functions at each clique are the conditional probability matrices  $P(R|S)$ .

Thus propagation is complete and any further  $\lambda$  messages will contain no evidence until another variable is instantiated

## Timetable for the near future:

Thurs 8 Feb 9.00 Lecture 11 Graphical Models (Sesh)

Thurs 8 Feb 10.00 Tutorial 6 Graphical Models

Tue 13 Feb 16.00: Lecture 12 (Marc)

Tue 13 Feb 17.00: Lecture 13 (Marc)

Thurs 15 Feb 09.00: Lecture 14 (Marc)

Thurs 15 Feb 10.00: Tutorial 5 (Duncan)