Multiple connected networks

3 classes of algorithms

(i) variable elimination
(ii) conditioning
(iii) junction tree

Variable elimination:

Calculate probability by marginalizing the joint distribution.
Makes use of independance property of BN.
Let
$X = \{X_1\ X_2 \cdots X_n\}$.

Posteriori probability of X_H given a subset of evidence X_E and the remaining variables are X_R, such that

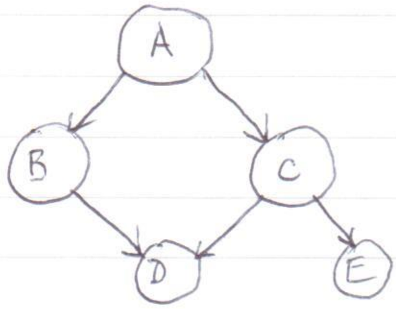
$$X = \{X_H \cup X_E \cup X_R\}$$

$$\therefore\ P(X_H | X_E) = P(X_H, X_E) / P(X_E).$$

$$P(X_H, X_E) = \sum_{X_R} P(X)$$

and $P(X_E) = \sum_{X_H} P(X_H, X_E)$ - (Probability of the evidence).

Consider the BN



We want $P(A|D)$.

$$P(A|D) = P(A,D)/P(D)$$

To calculate $P(A,D)$ we must eliminate B, C, E.

$$P(A,D) = \sum_B \sum_C \sum_E P(A)\,P(B|A)\,P(C|A)\,P(D|B,C)\,P(E|C).$$

By distributing the summations

$$P(A,D) = P(A) \sum_B P(B|A) \sum_C P(C|A)\,P(D|B,C) \sum_E P(E|C)$$

If variables are binary, then this implies a reduction from 32 operations to 9 operations.

Again consider

$P(c)$

c_1	c_2
0.8	0.2

$P(E|c)$

	c_1	c_2
E_1	0.9	0.7
E_2	0.1	0.3

$P(D|E)$

	E_1	E_2
D_1	0.7	0.4
D_2	0.3	0.6

$P(F|E)$

	E_1	E_2
F_1	0.9	0.5
F_2	0.1	0.5

We want to obtain

$$P(E|F=f_1) = P(E, F=f_1)/P(F=f_1)$$

Now, the joint probability

$$P(C,E,F,D) = P(C)\,P(E|C)\,P(F|E)\,P(D|E).$$

$P(C), P(E|C), P(F|E), P(D|E)$ given

$$P(E,F) = \sum_D P(F|E)\,P(D|E) \sum_C P(C)\,P(E|C)$$

We must do this for each E, given $F = f_1$:

$$P(e_1 f_1) = \sum_D P(f_1|e_1)\,P(D|e_1) \sum_C P(C)\,P(e_1|C)$$

$$= [0.9 \times 0.7 + 0.9 \times 0.3][0.9 \times 0.8 + 0.7 \times 0.2]$$

$$= [0.9][0.86] = 0.774$$

Similarly, we obtain $P(e_2, f_1)$.

Then,

$$P(f_1) = \sum_E P(E | f_1).$$

Finally, $P(e_1 | f_1) = P(e_1, f_1) / P(f_1)$

and $P(e_2 | f_1) = P(e_2, f_1) / P(f_1)$.

Conditioning

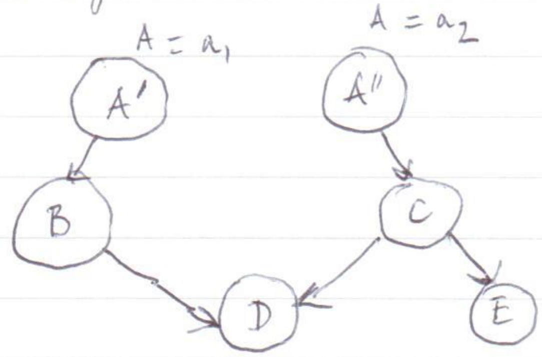
- An instantiated variable blocks the propagation of evidence in BN.
- <u>Cut</u> the graph at an instantiated variable and transform a multiconnected graph to a polytree.

Say, Probability of B given evidence E, conditioning on variable A?

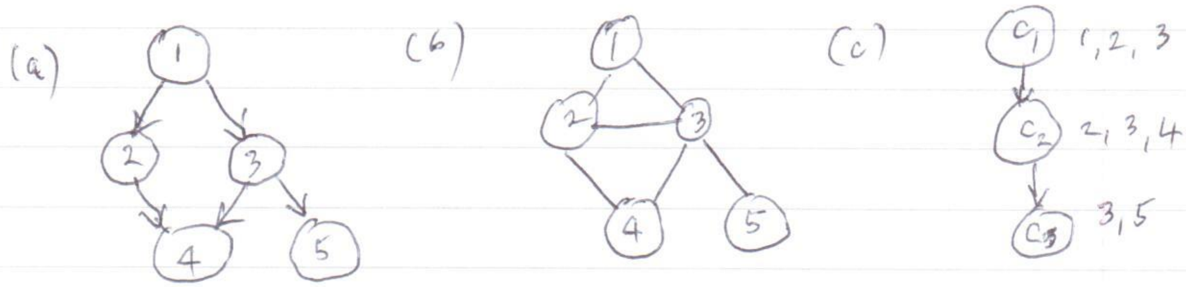
$$P(B|E) = \sum_i \underbrace{P(B | E, a_i)}_{\text{Posteriori probability obtained by propagation for each value of A.}} \underbrace{P(a_i | E)}_{\text{Weight}} \quad \text{(Total probability)}$$

Bayes rule: $P(a_i|E) = \alpha \underbrace{P(a_i)}_{\text{Propagated without evidence}} P(E|a_i)$.

BN transformed into a single connected network by instantiating A.



Junction tree algorithm:
Transformation of the BN to a junction tree, where each node is a group or cluster of variables.



Consider the BN with $clq_1 = \{1, 2, 3\}$, $clq_2 = \{2, 3, 4\}$ $clq_3 = \{3, 5\}$.

C: $C_1 = \{1, 2, 3\}$, $C_2 = \{2, 3, 4\}$, $C_3 = \{3, 5\}$

S: $S_1 = \phi$, $S_2 = \{2, 3\}$, $S_3 = \{3\}$.

R: $R_1 = \{1, 2, 3\}$, $R_2 = \{4\}$, $R_3 = \{5\}$; $R_i = C_i - S_i$

Potentials: $\psi(clq_1) = P(1)\, P(2|1)\, P(3|1)$, $\psi(clq_2) = P(4|3,2)$

$\psi(clq_3) = P(5|3)$.

The propagation proceeds in a similar way to belief propagation.

Bottom up:

1. Calculate λ message to send to parent clique $\lambda(C_i) = \sum_{R} \psi(C_i)$

2. Update the potential of each clique with λ messages from children $\psi(C_j)' = \lambda(C_i)\,\psi(C_j)$

3. Repeat the above two until the root clique is reached.
4. When reaching the root node obtain $P'(C_r) = \psi(C_r)'$.

Top-down:

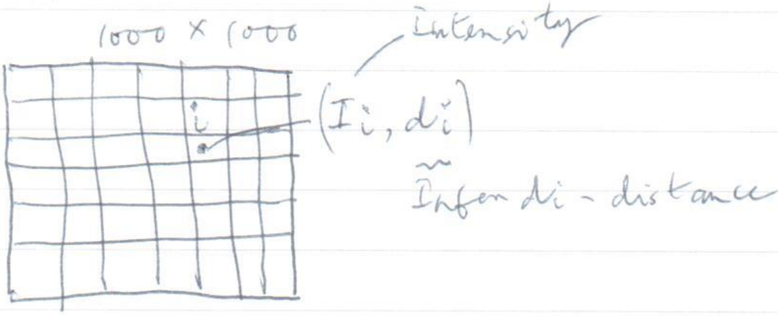
1. Calculate the π messages to send to each child node i:
   $\pi(C_i) = \sum_{C_j - S_i} P'(C_j)$.
2. Update the potential of each clique when receiving the π message of its parent:
   $P'(C_i) = \pi(C_i)\,\psi(C_i)'$.

3. Repeat the previous two steps until reaching the leaf nodes in the junction tree.

Pairwise Markov Random Fields

BP - used as an engine for low level computer vision problems (Freeman, Pasztor and Carmichael 2000).

Pairwise Markov Random Fields (MRF's) are attractive models.



or high resolution details / flow of images (optical)

observe y_i, infer x_i (some other quantity)

Statistical dependence between x_i and y_i.

Joint probability $\phi(x_i, y_i)$ - also called the "evidence" for x_i.

'structure' of the scene - context

No structure, leads to "ill-posed" problems.

'x_i should be compatible with nearby scene x_j'

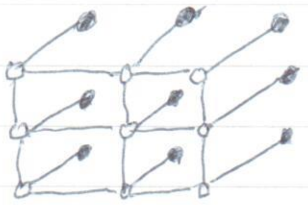
expressed by compatibility function $\psi_{ij}(x_i, x_j)$.

Joint probability of a scene x_i and an image y_i is given by

$$p(\{x\}, \{y\}) = \frac{1}{Z} \prod_{(ij)} \psi_{ij}(x_i, x_j) \prod_i \phi_i(x_i, y_i)$$

Z = normalization constant

Graphical representation



A square lattice pairwise MRF.

• observed image nodes y_i

o "hidden" scene nodes x_i

"pairwise" because the compatibility function only depends on pairs of sites i and j.

Remarks:

1. Undirected (contrast to BN) – no causal relation between parent and child.

2. $\psi_{ij}(x_i, x_j)$ used instead conditional probability $p(x_i|x_j)$ ($\because$ undirected)

3. Direct computation of marginal probabilities would take exponential time, hence need faster algorithm like BP.

We want to compute beliefs $b(x_i)$ for all i.

The Potts and Ising models

How MRF can be brought into a form of Potts model (Baxter 1982) recognizable to the physicists?

Let us define "interaction" between two neighbouring particles i and j at the two nodes by

$$J_{ij}(x_i, x_j) = \ln \psi_{ij}(x_i, x_j)$$

and the field at node i by

$$h_i(x_i) = \ln \phi(x_i, y_i).$$

($\because$ we are infering on y_i, we consider y_i as fixed and hence $h_i(x_i)$ subsumes y_i).

Potts model energy

$$E(\{x\}) = -\sum_{ij} J_{ij}(x_i, x_j) - \sum_i h_i(x_i)$$

Apply Boltzmann's law from statistical mechanics

$$p(\{x\}) = \frac{1}{Z} e^{-E(\{x_i\})/T}$$

We see that our pairwise MRF corresponds exactly to a Potts model with $T = 1$. (T = temperature)
Z = partition function.

If the number of states in each node is 2, then we get the "Ising model".

Physicists sometimes change variables from x_i to s_i, which can take on values of +1 or −1; and J_{ij} interactions have a symmetric form which can be written as a "spin glass" energy function (Mezard, Parisi and Virasoro 1987)

$$E(\{s\}) = - \sum_{ij} J_{ij} s_i s_j - \sum_i h_i s_i$$

In the context of Ising model, the inference problem of computing beliefs $b(x_i)$ can be mapped onto the physics problem of computing local "magnetization"

$$m_i \equiv b(s_i = 1) - b(s_i = -1).$$