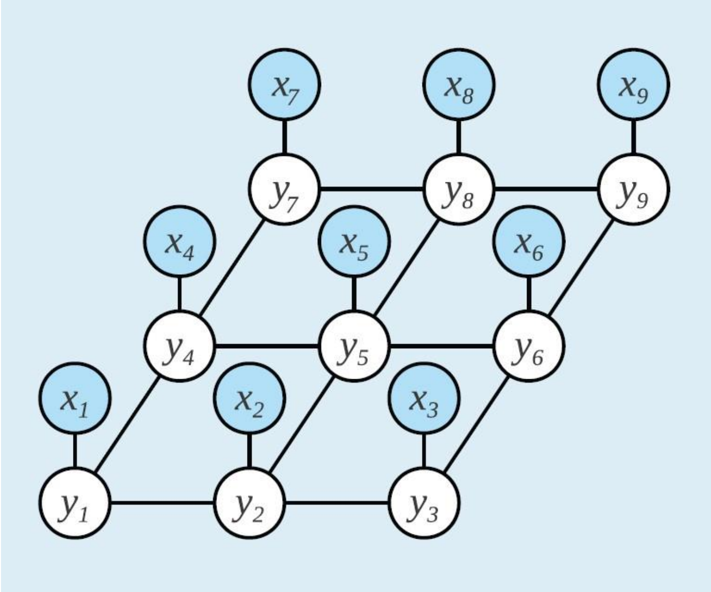
Probabilistic Graphical Models in Bioinformatics

Lecture 11: Undirected graphical models; Variable elimination



Hesam Montazeri
Department of Bioinformatics, IBB, University of Tehran

Ordibehesht 25, 1398
(May 15, 2019)

Undirected graphical models

- Bayesian networks or directed graphical models
  - Useful in many types of real-world domains
- Undirected graphical models
  - When there is no natural directionality to ascribe to the interaction between variables.
  - Often simpler framework in comparison to directed models in terms of the **independence structure** and the **inference task**.

- We restrict our attention to distributions over discrete space.

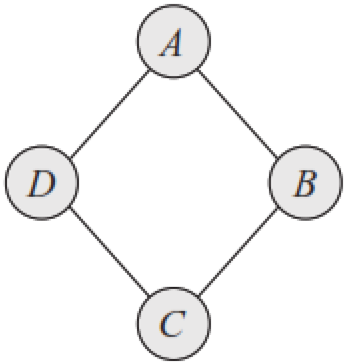
The *misconception* example

1. Four students study in pairs Alice and Bob; Bob and Charles; Charles and Debbie; Debbie and Alice.

2. Professor accidentally misspoke in the class, leading to a possible misconception

3. Students may have figured out the problem by reading the textbook

4. Students transmits their understanding to their partners

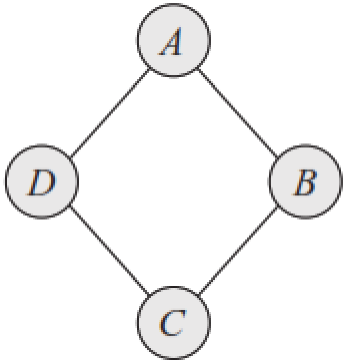


Study pairs for students

The *misconception* Example

- We need a model that satisfy the following independencies, but no other independencies

$$A \perp C \mid \{B, D\} \text{ and } B \perp D \mid \{A, C\}$$



- Drawbacks of Bayesian networks in this example
  - Previously, we saw BN fails to capture the independence structure implied in this example.
  - Interaction between variables are symmetric.
- Markov network or undirected graphical models
  - Node: random variables
  - Edge: direct probabilistic interaction
- Question: How to parameterize undirected graphs?
  - We cannot use a standard CPD.
  - We need a symmetric parameterization.

Factors

- Let D be a set of random variables. A factor ϕ is defined to be a function from $Val(D)$ to *R*.

$\phi_1(A, B)$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

- A factor is nonnegative if all its entries are nonnegative. The set of variables *D* is called the scope of the factor and denoted *Scope*(ϕ).

- Factor is not necessarily normalized to one or even in [0, 1].

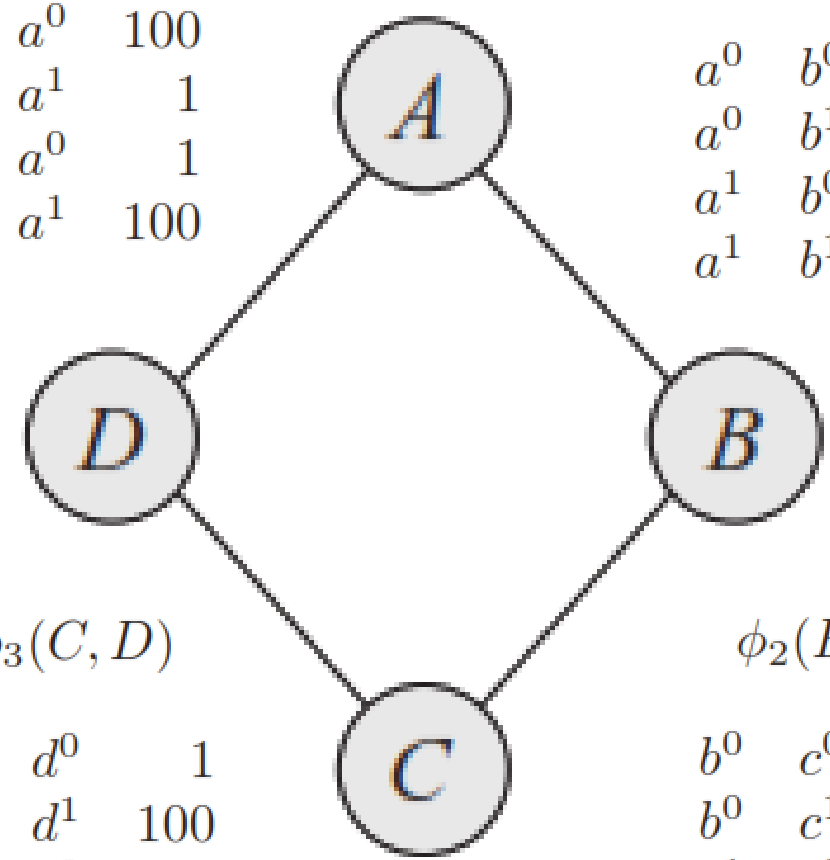
- A factor is aimed to capture affinity between related variables.

$\phi_4(D, A)$

d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10



$\phi_3(C, D)$

c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

For a^1, b^1, c^0, d^1

$$\phi_1(a^1, b^1) \cdot \phi_2(b^1, c^0) \cdot \phi_3(c^0, d^1) \cdot \phi_4(d^1, a^1) = 10 \cdot 1 \cdot 100 \cdot 100 = 100,000.$$

Joint distribution for Misconception example

$$P(a,b,c,d) = \frac{1}{Z}\phi_1(a,b).\,\phi_2(b,c).\,\phi_3(c,d).\,\phi_4(d,a)$$

where

$$Z = \sum_{a,b,c,d} \phi_1(a,b).\,\phi_2(b,c).\,\phi_3(c,d).\,\phi_4(d,a)$$

Assignment				Unnormalized	Normalized
a^0	b^0	c^0	d^0	300,000	0.04
a^0	b^0	c^0	d^1	300,000	0.04
a^0	b^0	c^1	d^0	300,000	0.04
a^0	b^0	c^1	d^1	30	$4.1 \cdot 10^{-6}$
a^0	b^1	c^0	d^0	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^0	d^1	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^1	d^0	5,000,000	0.69
a^0	b^1	c^1	d^1	500	$6.9 \cdot 10^{-5}$
a^1	b^0	c^0	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^0	d^1	1,000,000	0.14
a^1	b^0	c^1	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^1	d^1	100	$1.4 \cdot 10^{-5}$
a^1	b^1	c^0	d^0	10	$1.4 \cdot 10^{-6}$
a^1	b^1	c^0	d^1	100,000	0.014
a^1	b^1	c^1	d^0	100,000	0.014
a^1	b^1	c^1	d^1	100,000	0.014

Answering queries:

- $P(b^1) \approx 0.732$: Bob is 73 percent likely to have the misconception.
- $P(b^1 \mid c^0) \approx 0.06$: if Charles does not have the misconception, Bob is less likely to have the misconception.

Great flexibility in representing interactions

- If we want to change the nature of interactions between A and B, we can simply modify the entries in the factor without having to deal with normalization constraints

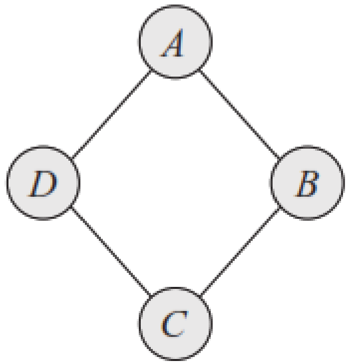
- The flip side of this flexibility is that the effects of these changes are not always intuitively understandable.

Factorization and independence

- As in Bayesian networks, there is a tight connection between the factorization of the distribution and its independence properties.

- $P: X \perp Y \mid Z$ if and only if we can write P in the form $P(X,Y,Z) = \phi_1(X,Z)\phi_2(Y,Z)$.

$$P(A,B,C,D) = \left[\frac{1}{Z}\phi_1(A,B)\phi_2(B,C)\right]\phi_3(C,D)\phi_4(A,D) \quad \Longleftrightarrow \quad (B \perp D \mid A, C).$$



- From the graph: B and D are separated given A and C.

- Independence properties of distribution P correspond directly to separation properties in the graph

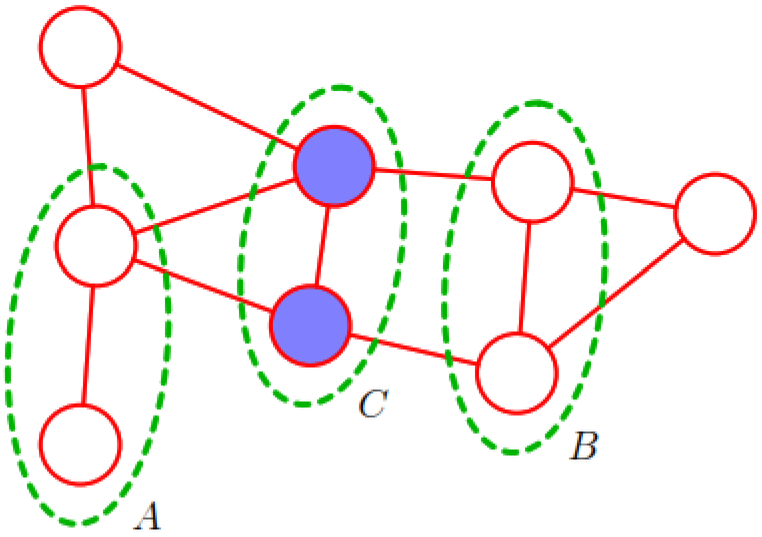
Markov network independencies

University of TEHRAN

- As in the case of Bayesian networks, a Markov network also encodes a set of independence assumptions.

- Separation algorithm is trivial for Markov networks.

- In Markov networks, a path $X_1 - X_2 - \cdots - X_k$ is active given Z if none of the X_i's is in Z.

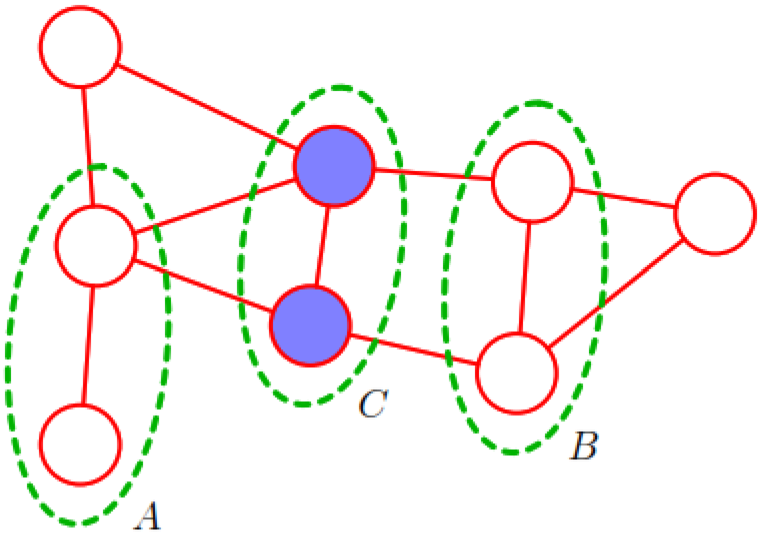


$$A \perp\!\!\!\perp B \mid C$$

Markov network independencies

We say that a set of nodes $\boldsymbol{Z}$ separates $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *in* $\mathcal{H}$*, denoted* $\mathrm{sep}_{\mathcal{H}}(\boldsymbol{X};\boldsymbol{Y} \mid \boldsymbol{Z})$*, if there is no active path between any node* $X \in \boldsymbol{X}$ *and* $Y \in \boldsymbol{Y}$ *given* $\boldsymbol{Z}$*. We define the* global independencies *associated with* $\mathcal{H}$ *to be:*

$$\mathcal{I}(\mathcal{H}) = \{(\boldsymbol{X} \perp \boldsymbol{Y} \mid \boldsymbol{Z}) \ : \ \mathrm{sep}_{\mathcal{H}}(\boldsymbol{X};\boldsymbol{Y} \mid \boldsymbol{Z})\}. \qquad \blacksquare$$

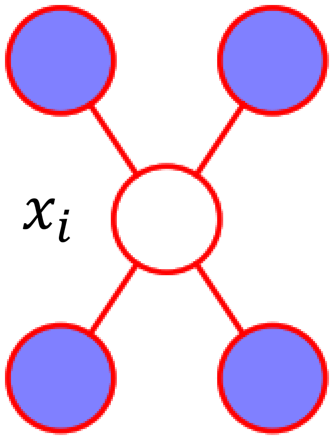


$A \perp\!\!\!\perp B \mid C$

Markov blanket

- For Markov networks, the Markov blanket of a node x_i consists of the set of neighboring nodes.

- Conditional distribution of x_i, conditioned on all remaining variables in the graph, is dependent only on the variable in the Markov blanket.



More on parametrization

- The parametrization of Markov networks is not as intuitive as that of Bayesian networks
  - Factors do not correspond either to probabilities or to CPDs.
  - Hard to elicit from people
  - Hard to learn from data

- We parameterize the graph by associating a set of factors with it.
  - **One obvious idea:** associate parameters directly with edges in the graph.
    - Insufficient to parameterize a full distribution (See Example 4.1 of the textbook)
  - Allowing factors over arbitrary subsets of variables

Factor product

$\phi_1(A, B)$

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

$\phi_2(B, C)$

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

$$\psi(A, B, C) = \phi_1(A, B).\phi_2(B, C)$$

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

Two factors ϕ_1 and ϕ_2 are multiplied in a way that matches-up the common part B

Reduced Markov networks

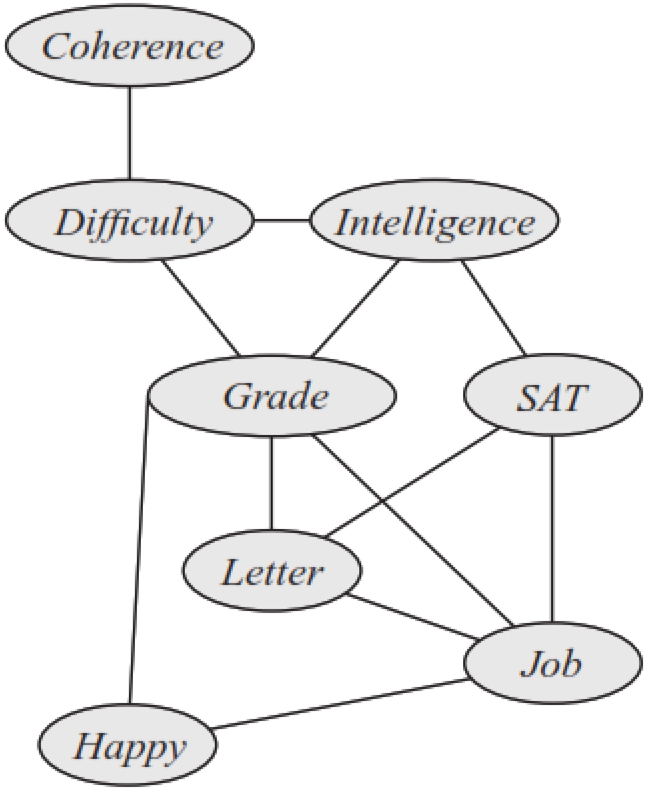
- Conditioning a distribution corresponds to all entries in the joint distribution that are consistent with the event U=u and renormalizing the remaining entries to sum to 1.

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

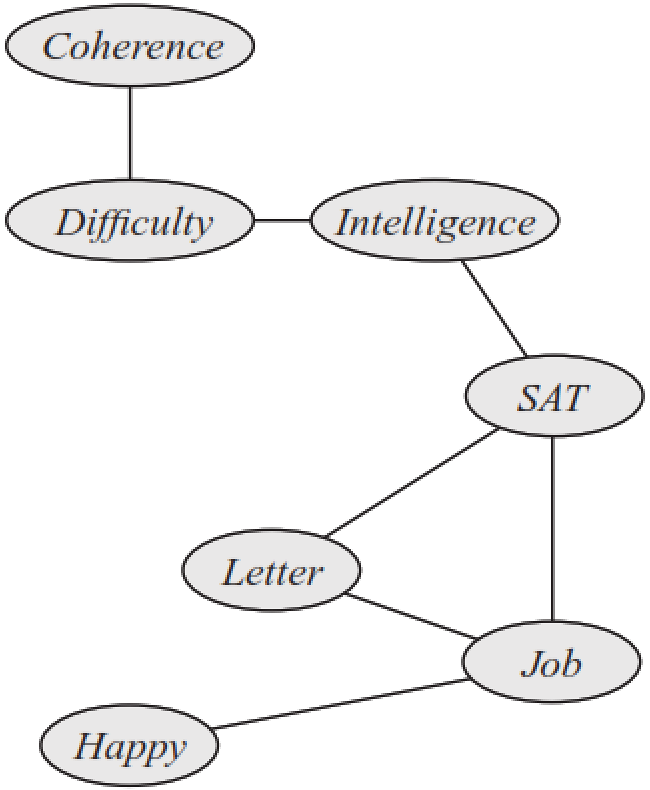
Reduced to the context $C = c^1$

a^1	b^1	c^1	0.25
a^1	b^2	c^1	0.08
a^2	b^1	c^1	0.05
a^2	b^2	c^1	0
a^3	b^1	c^1	0.15
a^3	b^2	c^1	0.09

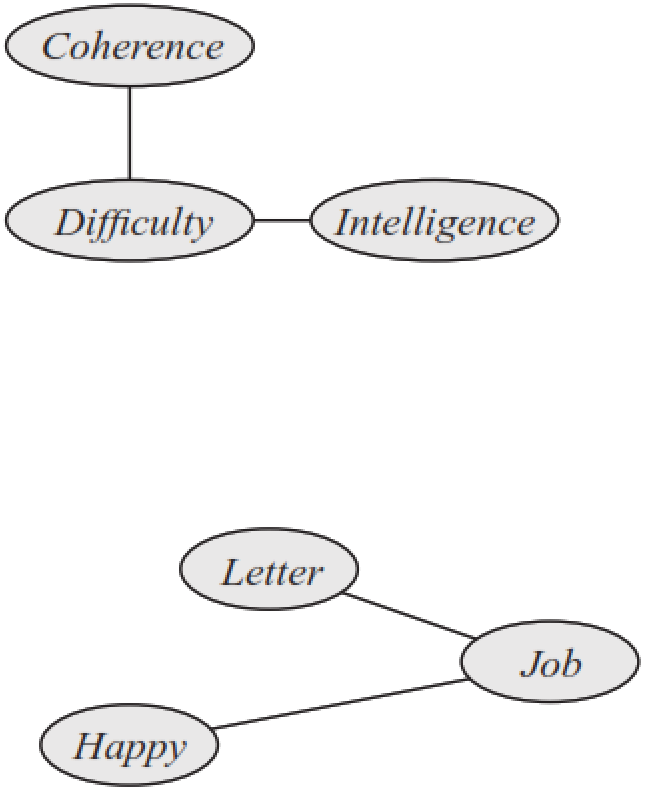
Reduced Markov networks- examples



The initial set of factors



Reduced to the context G=g



Reduced to the context G=g, S=s

Gibbs Distribution

- A distribution P_Φ is a Gibbs distribution parameterized by a set of factors $\Phi = \{\phi_1(D_1), \dots, \phi_K(D_K)\}$ if it is defined as follows:

$$P_\Phi(X_1, \dots, \mathrm{X_n}) = \frac{1}{Z}\tilde{P}_\Phi(X_1, \dots, \mathrm{X_n})$$

where

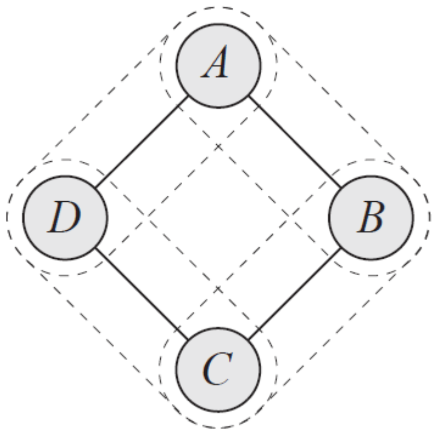
$$\tilde{P}_\Phi(X_1, \dots, \mathrm{X_n}) = \phi_1(D_1) \times \phi_2(D_2) \times \cdots \times \phi_K(D_K)$$

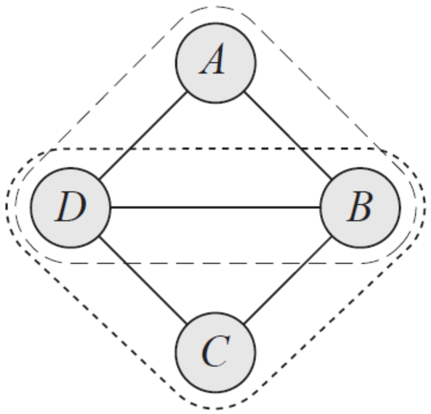
and Z is a normalizing constant called the partition function.

$$Z = \sum_{X_1, \dots, X_n} \tilde{P}_\Phi(X_1, \dots, \mathrm{X_n})$$

Markov network factorization

- We say that a distribution P_Φ with $\Phi = \{\phi_1(D_1), \ldots, \phi_K(D_K)\}$ factorizes over a Markov network H if each D_K is a complete subgraph of H.
- The factors that parameterize a Marko network are called *clique potentials*.





- Factors are not equivalent to marginal probabilities of the variables

Marginal distribution over A, B

a^0	b^0	0.13
a^0	b^1	0.69
a^1	b^0	0.14
a^1	b^1	0.04

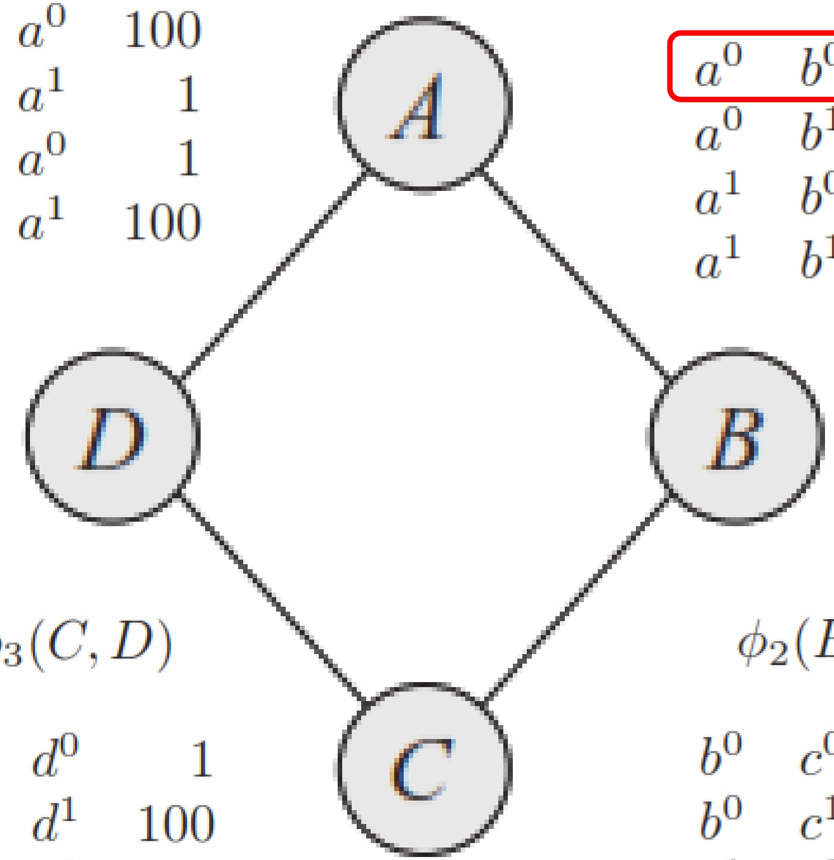
- A factor is only one contribution to the overall joint distribution

$\phi_4(D, A)$

d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1(A, B)$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10



$\phi_3(C, D)$

c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2(B, C)$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

Pairwise Markov networks

- Pairwise Markov networks arise in many contexts.
- Representing distributions where all of the factors are over single variables or pairs of variables.
- A commonly used class of pairwise Markov networks structured in the form of a grid

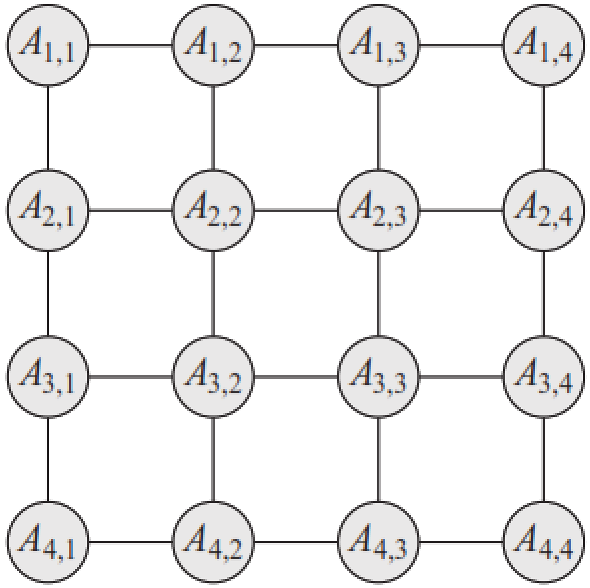
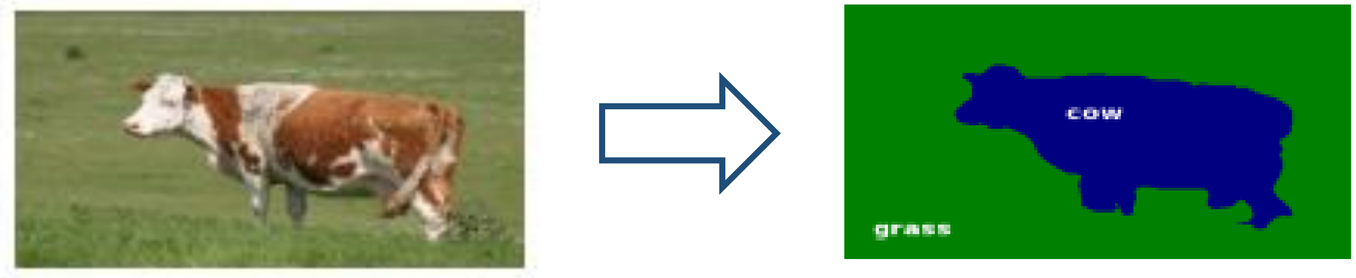


Image segmentation



Inference

Probabilistic inference

- We want to answer queries that may involve evidence, $P(Y \mid E = e)$.

$$P(\,Y \mid E = e\,) = \frac{P(Y, e)}{P(e)} = \frac{P(Y, e)}{\sum_y P(Y = y, e)}$$

- Often we only query a subset Y of all domain variables. We need to marginalize over the remaining variables

$$P(Y, e) = \sum_z P(Y\,, e, Z = z)$$

Inference algorithms

- Exact inference
  - Variable elimination
  - Clique trees

- Approximate inference
  - Optimization
  - Particle-based inference (sampling)

Variable Elimination: the basic ideas

- The BN structure allows use of dynamic programming techniques to perform inference for certain large and complex networks
  - Reading assignment: A.3.3

- Consider a simple network $A \rightarrow B \rightarrow C \rightarrow D$

- Question: compute $P(B)$. How many arithmetic operations are required?

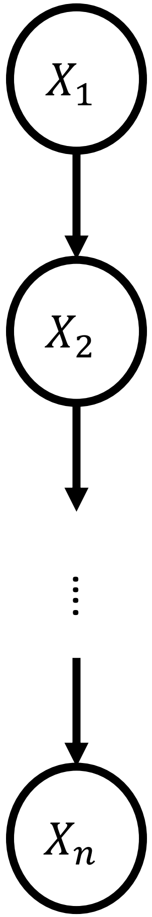
$$P(B) = \sum_a P(a)P(B \mid a)$$

The algorithm does not compute single values but rather sets of values.

- Similarly

$$P(C) = \sum_b P(b)P(C \mid b)$$

Example: inference in chains



- Consider a chain with n variables, each with k values.
- The algorithm computes $P(X_{i+1})$ from $P(X_i)$

$$P(X_{i+1}) = \sum_{x_i} P(X_{i+1} \mid x_i) P(x_i)$$

- What is the complexity of the inference algorithm?

Goal: compute P(D) for the chain $A \rightarrow B \rightarrow C \rightarrow D$ with binary variables

$P(d^1)$

$$\begin{aligned}
& P(a^1) \; P(b^1 \mid a^1) \; P(c^1 \mid b^1) \; P(d^1 \mid c^1) \\
+\; & P(a^2) \; P(b^1 \mid a^2) \; P(c^1 \mid b^1) \; P(d^1 \mid c^1) \\
+\; & P(a^1) \; P(b^2 \mid a^1) \; P(c^1 \mid b^2) \; P(d^1 \mid c^1) \\
+\; & P(a^2) \; P(b^2 \mid a^2) \; P(c^1 \mid b^2) \; P(d^1 \mid c^1) \\
+\; & P(a^1) \; P(b^1 \mid a^1) \; P(c^2 \mid b^1) \; P(d^1 \mid c^2) \\
+\; & P(a^2) \; P(b^1 \mid a^2) \; P(c^2 \mid b^1) \; P(d^1 \mid c^2) \\
+\; & P(a^1) \; P(b^2 \mid a^1) \; P(c^2 \mid b^2) \; P(d^1 \mid c^2) \\
+\; & P(a^2) \; P(b^2 \mid a^2) \; P(c^2 \mid b^2) \; P(d^1 \mid c^2)
\end{aligned}$$

$P(d^2)$

$$\begin{aligned}
& P(a^1) \; P(b^1 \mid a^1) \; P(c^1 \mid b^1) \; P(d^2 \mid c^1) \\
+\; & P(a^2) \; P(b^1 \mid a^2) \; P(c^1 \mid b^1) \; P(d^2 \mid c^1) \\
+\; & P(a^1) \; P(b^2 \mid a^1) \; P(c^1 \mid b^2) \; P(d^2 \mid c^1) \\
+\; & P(a^2) \; P(b^2 \mid a^2) \; P(c^1 \mid b^2) \; P(d^2 \mid c^1) \\
+\; & P(a^1) \; P(b^1 \mid a^1) \; P(c^2 \mid b^1) \; P(d^2 \mid c^2) \\
+\; & P(a^2) \; P(b^1 \mid a^2) \; P(c^2 \mid b^1) \; P(d^2 \mid c^2) \\
+\; & P(a^1) \; P(b^2 \mid a^1) \; P(c^2 \mid b^2) \; P(d^2 \mid c^2) \\
+\; & P(a^2) \; P(b^2 \mid a^2) \; P(c^2 \mid b^2) \; P(d^2 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& (P(a^1)P(b^1 \mid a^1) + P(a^2)P(b^1 \mid a^2)) \; P(c^1 \mid b^1) \; P(d^1 \mid c^1) \\
+\; & (P(a^1)P(b^2 \mid a^1) + P(a^2)P(b^2 \mid a^2)) \; P(c^1 \mid b^2) \; P(d^1 \mid c^1) \\
+\; & (P(a^1)P(b^1 \mid a^1) + P(a^2)P(b^1 \mid a^2)) \; P(c^2 \mid b^1) \; P(d^1 \mid c^2) \\
+\; & (P(a^1)P(b^2 \mid a^1) + P(a^2)P(b^2 \mid a^2)) \; P(c^2 \mid b^2) \; P(d^1 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& (P(a^1)P(b^1 \mid a^1) + P(a^2)P(b^1 \mid a^2)) \; P(c^1 \mid b^1) \; P(d^2 \mid c^1) \\
+\; & (P(a^1)P(b^2 \mid a^1) + P(a^2)P(b^2 \mid a^2)) \; P(c^1 \mid b^2) \; P(d^2 \mid c^1) \\
+\; & (P(a^1)P(b^1 \mid a^1) + P(a^2)P(b^1 \mid a^2)) \; P(c^2 \mid b^1) \; P(d^2 \mid c^2) \\
+\; & (P(a^1)P(b^2 \mid a^1) + P(a^2)P(b^2 \mid a^2)) \; P(c^2 \mid b^2) \; P(d^2 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& \tau_1(b^1) \; P(c^1 \mid b^1) \; P(d^1 \mid c^1) \\
+\; & \tau_1(b^2) \; P(c^1 \mid b^2) \; P(d^1 \mid c^1) \\
+\; & \tau_1(b^1) \; P(c^2 \mid b^1) \; P(d^1 \mid c^2) \\
+\; & \tau_1(b^2) \; P(c^2 \mid b^2) \; P(d^1 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& \tau_1(b^1) \; P(c^1 \mid b^1) \; P(d^2 \mid c^1) \\
+\; & \tau_1(b^2) \; P(c^1 \mid b^2) \; P(d^2 \mid c^1) \\
+\; & \tau_1(b^1) \; P(c^2 \mid b^1) \; P(d^2 \mid c^2) \\
+\; & \tau_1(b^2) \; P(c^2 \mid b^2) \; P(d^2 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& (\tau_1(b^1)P(c^1 \mid b^1) + \tau_1(b^2)P(c^1 \mid b^2)) \; P(d^1 \mid c^1) \\
+\; & (\tau_1(b^1)P(c^2 \mid b^1) + \tau_1(b^2)P(c^2 \mid b^2)) \; P(d^1 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& (\tau_1(b^1)P(c^1 \mid b^1) + \tau_1(b^2)P(c^1 \mid b^2)) \; P(d^2 \mid c^1) \\
+\; & (\tau_1(b^1)P(c^2 \mid b^1) + \tau_1(b^2)P(c^2 \mid b^2)) \; P(d^2 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& \tau_2(c^1) \; P(d^1 \mid c^1) \\
+\; & \tau_2(c^2) \; P(d^1 \mid c^2)
\end{aligned}$$

$$\begin{aligned}
& \tau_2(c^1) \; P(d^2 \mid c^1) \\
+\; & \tau_2(c^2) \; P(d^2 \mid c^2)
\end{aligned}$$

Total operations: 12 multiplications and 6 additions

four multiplications and two additions for $\tau_1(B)$; similarly for $\tau_2(C)$ and $P(D)$

Naïve computation requires 16.3=48 multiplications and 14 additions

Summary of computation

- We performed the following steps:

$$P(D) = \sum_C \sum_B \sum_A P(A)P(B \mid A)P(C \mid B)P(D \mid C).$$

- By pushing in the summations we obtained

$$\sum_C P(D \mid C) \sum_B P(C \mid B) \sum_A \underbrace{P(A)P(B \mid A)}_{\psi_1(A,B) = P(A)P(B \mid A)}.$$

$$\underbrace{\psi_1(A,B) = P(A)P(B \mid A)}_{\tau_1(B) = \sum_A \psi_1(A,B)}$$

For each *b*

$$\tau_1(b) = \sum_A \psi_1(A,b)$$

$$\begin{aligned} \psi_2(B,C) &= \tau_1(B)P(C \mid B) \\ \tau_2(C) &= \sum_B \psi_2(B,C). \end{aligned}$$

We then use $\tau_2(C)$ to compute P(D)

Factor marginalization

Summing out Y in ψ:

$$\psi(\boldsymbol{X}) = \sum_{Y} \phi(\boldsymbol{X}, Y).$$

a^1	b^1	c^1	0.25
a^1	b^1	c^2	0.35
a^1	b^2	c^1	0.08
a^1	b^2	c^2	0.16
a^2	b^1	c^1	0.05
a^2	b^1	c^2	0.07
a^2	b^2	c^1	0
a^2	b^2	c^2	0
a^3	b^1	c^1	0.15
a^3	b^1	c^2	0.21
a^3	b^2	c^1	0.09
a^3	b^2	c^2	0.18

a^1	c^1	0.33
a^1	c^2	0.51
a^2	c^1	0.05
a^2	c^2	0.07
a^3	c^1	0.24
a^3	c^2	0.39

Summing out *B*

Example revisited

- We write the joint distribution as

$$P(A, B, C, D) = \phi_A \cdot \phi_B \cdot \phi_C \cdot \phi_D.$$

- Marginal distribution over D is

$$P(D) = \sum_C \sum_B \sum_A P(A, B, C, D).$$

- We can conclude

$$\begin{aligned} P(D) &= \sum_C \sum_B \sum_A \phi_A \cdot \phi_B \cdot \phi_C \cdot \phi_D \\ &= \sum_C \sum_B \phi_C \cdot \phi_D \cdot \left(\sum_A \phi_A \cdot \phi_B \right) \\ &= \sum_C \phi_D \cdot \left(\sum_B \phi_C \cdot \left(\sum_A \phi_A \cdot \phi_B \right) \right), \end{aligned}$$

- In general the following task is called *sum-product* inference task

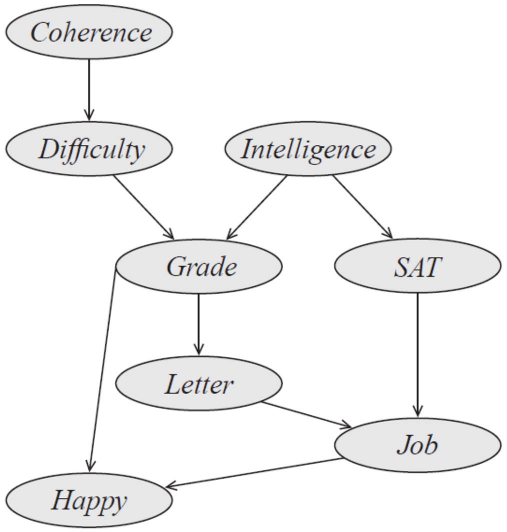
$$\sum_{\boldsymbol{Z}} \prod_{\phi \in \Phi} \phi.$$

Algorithm 9.1 Sum-product variable elimination algorithm

Procedure Sum-Product-VE (
  Φ, // Set of factors
  $\boldsymbol{Z}$, // Set of variables to be eliminated
  $\prec$ // Ordering on $\boldsymbol{Z}$
)
  Let $Z_1, \ldots, Z_k$ be an ordering of $\boldsymbol{Z}$ such that
    $Z_i \prec Z_j$ if and only if $i < j$
  for $i = 1, \ldots, k$
    $\Phi \leftarrow$ Sum-Product-Eliminate-Var(Φ, Z_i)
  $\phi^* \leftarrow \prod_{\phi \in \Phi} \phi$
  return ϕ^*

Procedure Sum-Product-Eliminate-Var (
  Φ, // Set of factors
  Z // Variable to be eliminated
)
  $\Phi' \leftarrow \{\phi \in \Phi \ : \ Z \in Scope[\phi]\}$
  $\Phi'' \leftarrow \Phi - \Phi'$
  $\psi \leftarrow \prod_{\phi \in \Phi'} \phi$
  $\tau \leftarrow \sum_Z \psi$
  return $\Phi'' \cup \{\tau\}$

$$
\begin{aligned}
P(C,D,I,G,S,L,J,H) &= P(C)P(D \mid C)P(I)P(G \mid I,D)P(S \mid I) \\
&\quad P(L \mid G)P(J \mid L,S)P(H \mid G,J) \\
&= \phi_C(C)\phi_D(D,C)\phi_I(I)\phi_G(G,I,D)\phi_S(S,I) \\
&\quad \phi_L(L,G)\phi_J(J,L,S)\phi_H(H,G,J).
\end{aligned}
$$



We apply the VE algorithm to compute P(J). We use the elimination ordering:
C, D, I, H, G, S, L

1. Eliminating *C*

$$
\begin{aligned}
\psi_1(C,D) &= \phi_C(C) \cdot \phi_D(D,C) \\
\tau_1(D) &= \sum_C \psi_1.
\end{aligned}
$$

2. Eliminating *D*

$$
\begin{aligned}
\psi_2(G,I,D) &= \phi_G(G,I,D) \cdot \tau_1(D) \\
\tau_2(G,I) &= \sum_D \psi_2(G,I,D).
\end{aligned}
$$

3. Eliminating *I*

$$
\begin{aligned}
\psi_3(G,I,S) &= \phi_I(I) \cdot \phi_S(S,I) \cdot \tau_2(G,I) \\
\tau_3(G,S) &= \sum_I \psi_3(G,I,S).
\end{aligned}
$$

4. Eliminating *H*

$$
\begin{aligned}
\psi_4(G,J,H) &= \phi_H(H,G,J) \\
\tau_4(G,J) &= \sum_H \psi_4(G,J,H).
\end{aligned}
$$

5. Eliminating *G*

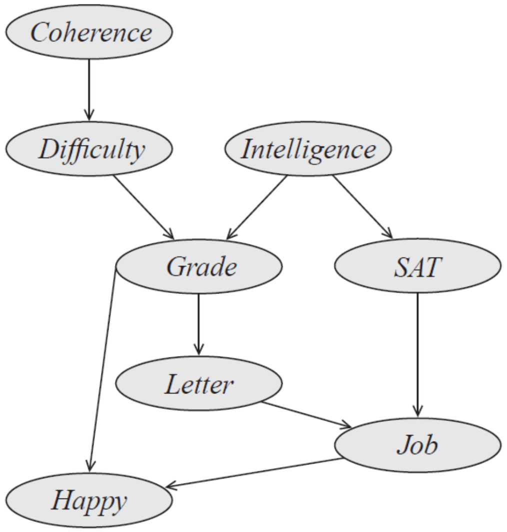
$$\psi_5(G,J,L,S) = \tau_4(G,J)\cdot\tau_3(G,S)\cdot\phi_L(L,G)$$
$$\tau_5(J,L,S) = \sum_G \psi_5(G,J,L,S).$$

6. Eliminating *S*

$$\psi_6(J,L,S) = \tau_5(J,L,S)\cdot\phi_J(J,L,S)$$
$$\tau_6(J,L) = \sum_S \psi_6(J,L,S).$$

6. Eliminating *L*

$$\psi_7(J,L) = \tau_6(J,L)$$
$$\tau_7(J) = \sum_L \psi_7(J,L).$$



Step	Variable eliminated	Factors used	Variables involved	New factor
1	C	$\phi_C(C)$, $\phi_D(D,C)$	C,D	$\tau_1(D)$
2	D	$\phi_G(G,I,D)$, $\tau_1(D)$	G,I,D	$\tau_2(G,I)$
3	I	$\phi_I(I)$, $\phi_S(S,I)$, $\tau_2(G,I)$	G,S,I	$\tau_3(G,S)$
4	H	$\phi_H(H,G,J)$	H,G,J	$\tau_4(G,J)$
5	G	$\tau_4(G,J)$, $\tau_3(G,S)$, $\phi_L(L,G)$	G,J,L,S	$\tau_5(J,L,S)$
6	S	$\tau_5(J,L,S)$, $\phi_J(J,L,S)$	J,L,S	$\tau_6(J,L)$
7	L	$\tau_6(J,L)$	J,L	$\tau_7(J)$