CS 3750 Machine Learning

Lecture 2

Graphical models

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square, x4-8845

http://www.cs.pitt.edu/~milos/courses/cs3750/

Graphical models

- **Represent complex multivariate probabilistic models**
  - **multivariate -> multiple random variables**

$$P(\mathbf{X}) = P(X_1, X_2, .., X_d)$$

$$p(\mathbf{X}) = p(X_1, X_2, .., X_d)$$

- **Parametric distribution models:**
  - Bernoulli (outcome of coin flip)
  - Binomial (outcome of multiple coin flips)
  - Multinomial (outcome of die)
  - Poisson
  - Exponential
  - Gamma distribution
  - Gaussian (this one is multivariate)

Challenges for modeling complex multivariate distributions

How to model/parameterize complex multivariate distributions $P(\mathbf{X})$ with a large number of variables?

One solution:

- Decompose the distribution. Reduce the number of parameters, using some form of independence.

Two graphical models:

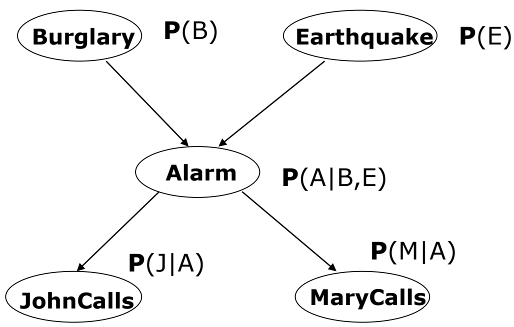
- **Bayesian belief networks (BBNs)**
- **Markov Random Fields (MRFs)**

- **Learning of these models** relies on the decomposition.

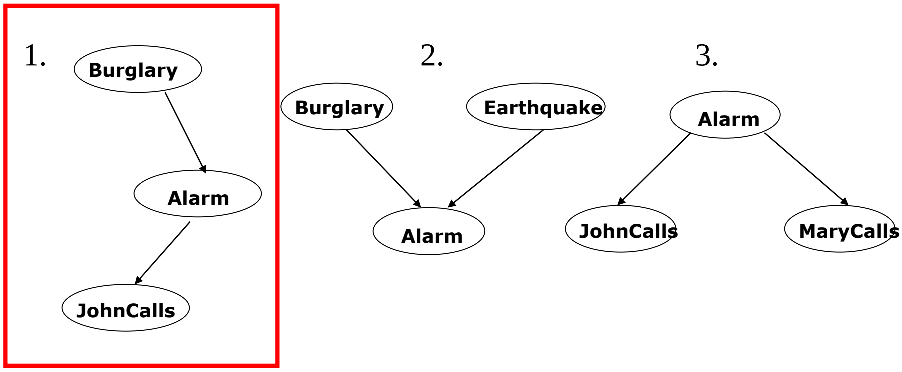
Bayesian belief network

Directed acyclic graph

- **Nodes** = random variables
- **Links** = direct (causal) dependencies
  Missing links encode different marginal and conditional independences



Graphical structure and independences

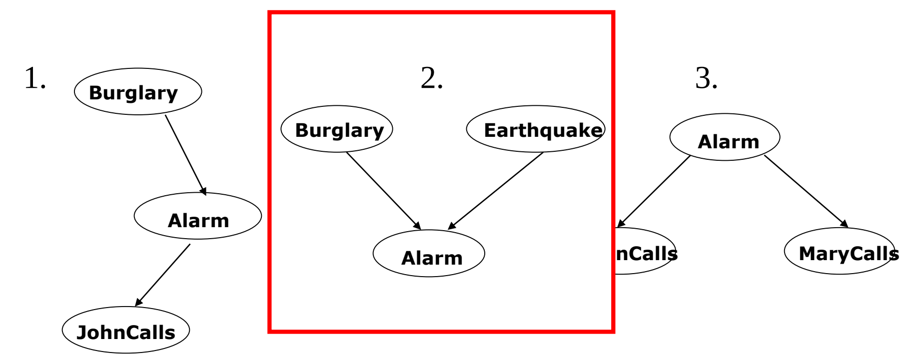


1. JohnCalls **is independent** of Burglary **given** Alarm

$$P(J \mid A, B) = P(J \mid A)$$

$$P(J, B \mid A) = P(J \mid A)P(B \mid A)$$

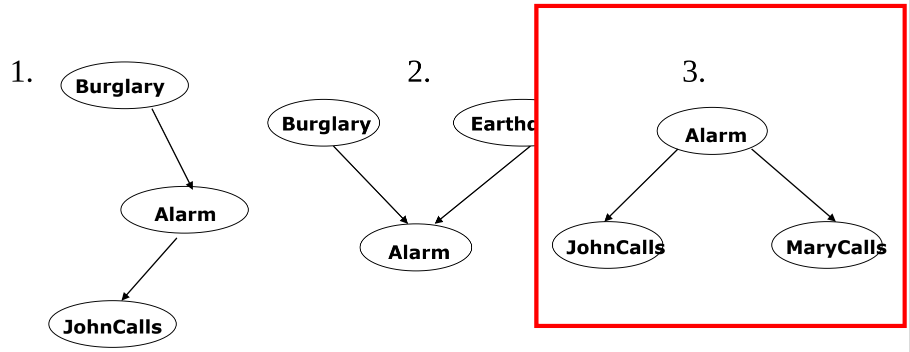
Independences in BBNs



2. Burglary **is independent** of Earthquake (not knowing Alarm)
   Burglary and Earthquake **become dependent** given Alarm !!

$$P(B, E) = P(B)P(E)$$

Independences in BBNs



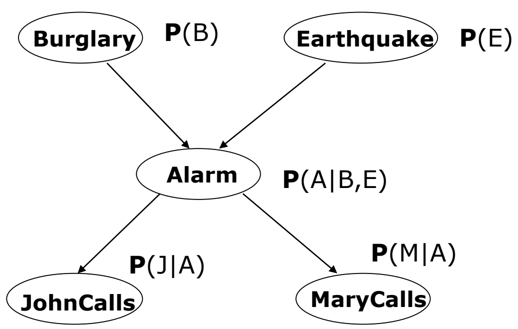
3. MaryCalls **is independent** of JohnCalls **given** Alarm

$$P(J \mid A, M) = P(J \mid A)$$

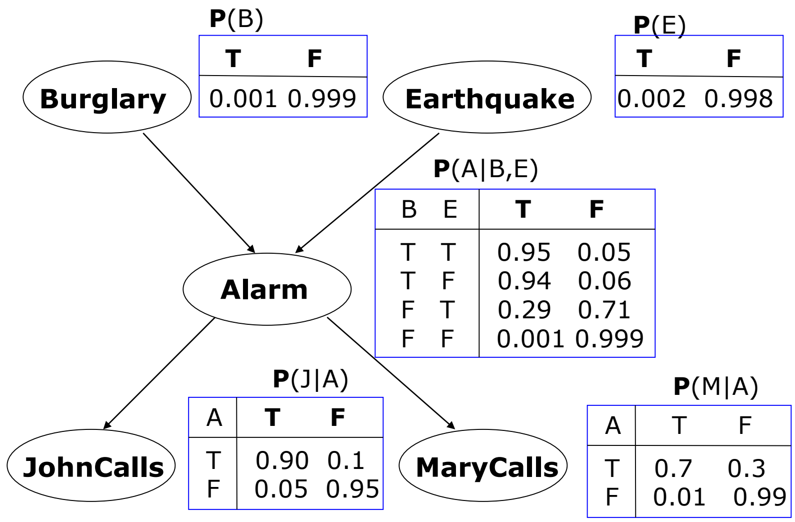
$$P(J, M \mid A) = P(J \mid A) P(M \mid A)$$

Bayesian belief network: parameters

2. **Local conditional distributions**
   - relate variables and their parents P(v | pa(v))



Bayesian belief network



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

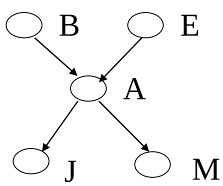
A	T	F
T	0.7	0.3
F	0.01	0.99

Full joint distribution in BBNs

The full joint distribution is defined as a product of local conditional distributions:

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

Example:



Assume the following assignment of values to random variables
$B = T, E = T, A = T, J = T, M = F$

Then its probability is:

$$P(B = T, E = T, A = T, J = T, M = F) =$$
$$P(B = T)P(E = T)P(A = T \mid B = T, E = T)P(J = T \mid A = T)P(M = F \mid A = T)$$

Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

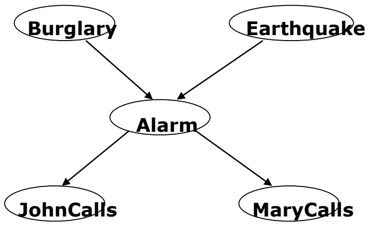
$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: binary (True, False) variables

# of parameters of the full joint:

?



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: binary (True, False) variables

# of parameters of the full joint:

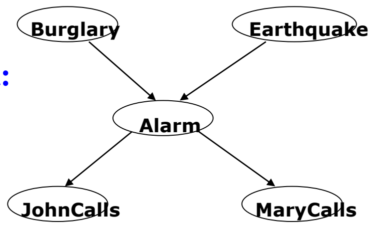
$$2^5 = 32$$

One parameter depends on the rest:

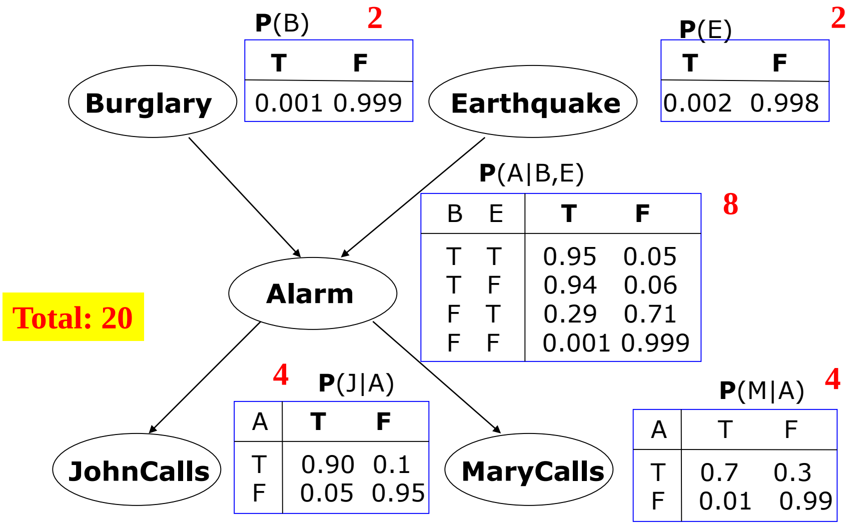
$$2^5 - 1 = 31$$

# of parameters of the BBN:

?



Bayesian belief network: parameters count



P(B) — 2

T	F
0.001	0.999

P(E) — 2

T	F
0.002	0.998

P(A|B,E) — 8

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A) — 4

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A) — 4

A	T	F
T	0.7	0.3
F	0.01	0.99

Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: 5 binary (True, False) variables

# of parameters of the full joint:

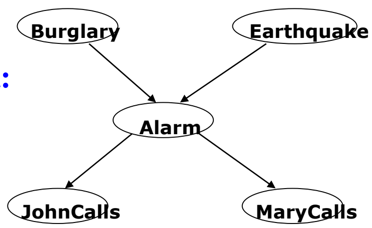
$$2^5 = 32$$

One parameter depends on the rest:

$$2^5 - 1 = 31$$

# of parameters of the BBN:

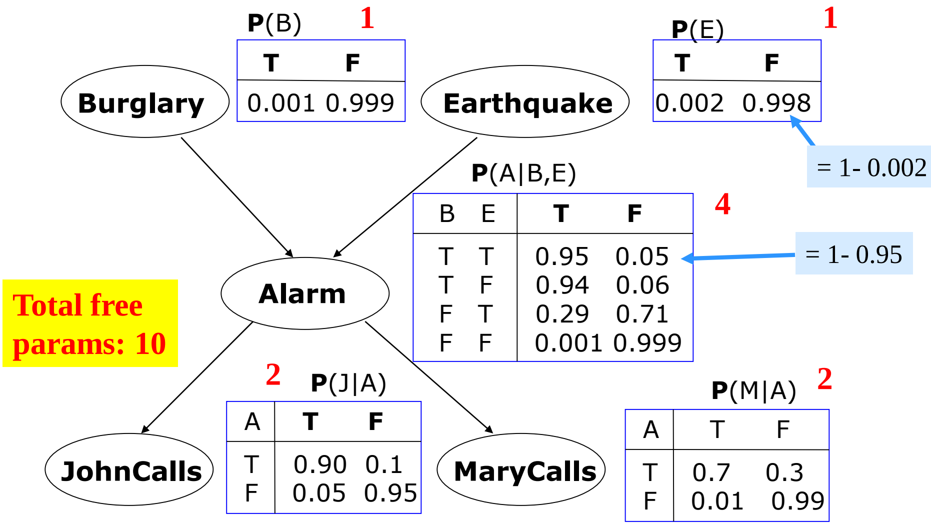
$$2^3 + 2(2^2) + 2(2) = 20$$



One parameter in every conditional depends on the rest:

?

Bayesian belief network: free parameters



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: 5 binary (True, False) variables

# of parameters of the full joint:

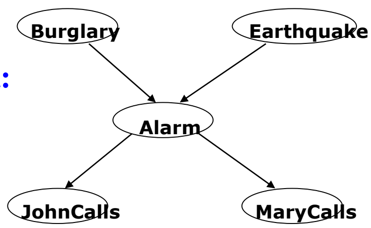
$$2^5 = 32$$

One parameter depends on the rest:

$$2^5 - 1 = 31$$

# of parameters of the BBN:

$$2^3 + 2(2^2) + 2(2) = 20$$

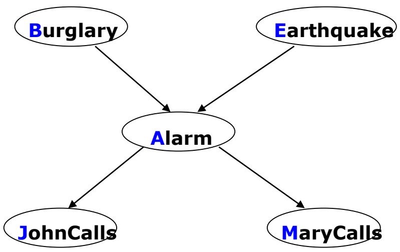


One parameter in every conditional depends on the rest:

$$2^2 + 2(2) + 2(1) = 10$$

Inference in Bayesian network

- **Bad news:**
  - Exact inference problem in BBNs is NP-hard (Cooper)
  - Approximate inference is NP-hard (Dagum, Luby)
- **But** very often we can achieve significant improvements
- Assume our Alarm network



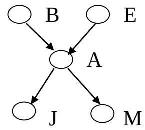
- Assume we want to compute: $P(J=T)$

Inference in Bayesian networks

- Full joint uses the decomposition
- **Calculation of marginals:**
  - Requires summation over variables we want to take out

$$P(J=T)=$$

$$=\sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F}P(B=b,E=e,A=a,J=T,M=m)$$



- How to compute sums and products more efficiently?

$$\sum_x af(x)=a\sum_x f(x)$$

Variable elimination

- **Variable elimination:**
  - E.g. Query $P(J=T)$ requires to eliminate A,B,E,M and this can be done in different order

$$P(J=T)=$$

$$=\sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T\mid A=a)P(M=m\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)$$

Variable elimination

Assume order: **M, E, B, A** to calculate $P(J=T)$

$$=\sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T\mid A=a)P(M=m\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)$$

Variable elimination

Assume order: **M**, **E, B**,**A** to calculate $P(J=T)$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J=T \mid A=a)P(M=m \mid A=a)P(A=a \mid B=b, E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b, E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m \mid A=a)\right]$$

Variable elimination

Assume order: **M, E, B, A** to calculate $P(J=T)$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J=T \mid A=a)P(M=m \mid A=a)P(A=a \mid B=b, E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b, E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m \mid A=a)\right]$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b, E=e)P(B=b)P(E=e) \; \boxed{1}$$

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T \mid A=a)P(M=m \mid A=a)P(A=a \mid B=b,E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b,E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m \mid A=a)\right]$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b,E=e)P(B=b)P(E=e) \quad 1$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T \mid A=a)P(B=b)\left[\sum_{e\in T,F} P(A=a \mid B=b,E=e)P(E=e)\right]$$

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T \mid A=a)P(M=m \mid A=a)P(A=a \mid B=b,E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b,E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m \mid A=a)\right]$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T \mid A=a)P(A=a \mid B=b,E=e)P(B=b)P(E=e) \quad 1$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T \mid A=a)P(B=b)\left[\sum_{e\in T,F} P(A=a \mid B=b,E=e)P(E=e)\right]$$

$\tau_1(A=a, B=b)$

$\tau_1(A=a, B=b) =$

	$A=T$	$A=F$	
$B=T$			$\sum_{e\in T,F} P(A=F \mid B=T, E=e)P(E=e)$
$B=F$			$\sum_{e\in T,F} P(A=F \mid B=F, E=e)P(E=e)$

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J=T \mid A=a) P(M=m \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e)$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e) \left[\sum_{m\in T,F} P(M=m \mid A=a) \right]$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e) \quad 1$$

$$= \sum_{a\in T,F} \sum_{b\in T,F} P(J=T \mid A=a) P(B=b) \left[\sum_{e\in T,F} P(A=a \mid B=b, E=e) P(E=e) \right]$$

$$= \sum_{a\in T,F} \sum_{b\in T,F} P(J=T \mid A=a) P(B=b)\, \tau_1(A=a, B=b)$$

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J=T \mid A=a) P(M=m \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e)$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e) \left[\sum_{m\in T,F} P(M=m \mid A=a) \right]$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e) \quad 1$$

$$= \sum_{a\in T,F} \sum_{b\in T,F} P(J=T \mid A=a) P(B=b) \left[\sum_{e\in T,F} P(A=a \mid B=b, E=e) P(E=e) \right]$$

$$= \sum_{a\in T,F} \sum_{b\in T,F} P(J=T \mid A=a) P(B=b)\, \tau_1(A=a, B=b)$$

$$= \sum_{a\in T,F} P(J=T \mid A=a) \left[\sum_{b\in T,F} P(B=b)\, \tau_1(A=a, B=b) \right]$$

Variable elimination

Assume order: **M, E, B, A** to calculate $P(J=T)$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T\mid A=a)P(M=m\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m\mid A=a)\right]$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)\quad 1$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T\mid A=a)P(B=b)\left[\sum_{e\in T,F} P(A=a\mid B=b,E=e)P(E=e)\right]$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T\mid A=a)P(B=b)\,\tau_1(A=a,B=b)$$

$$= \sum_{a\in T,F} P(J=T\mid A=a)\left[\sum_{b\in T,F} P(B=b)\,\tau_1(A=a,B=b)\right]$$

$$\tau_2(A=a)$$

Variable elimination

Assume order: **M, E, B, A** to calculate $P(J=T)$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F}\sum_{m\in T,F} P(J=T\mid A=a)P(M=m\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)\left[\sum_{m\in T,F} P(M=m\mid A=a)\right]$$

$$= \sum_{b\in T,F}\sum_{e\in T,F}\sum_{a\in T,F} P(J=T\mid A=a)P(A=a\mid B=b,E=e)P(B=b)P(E=e)\quad 1$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T\mid A=a)P(B=b)\left[\sum_{e\in T,F} P(A=a\mid B=b,E=e)P(E=e)\right]$$

$$= \sum_{a\in T,F}\sum_{b\in T,F} P(J=T\mid A=a)P(B=b)\,\tau_1(A=a,B=b)$$

$$= \sum_{a\in T,F} P(J=T\mid A=a)\left[\sum_{e\in T,F} P(B=b)\,\tau_1(A=a,B=b)\right]$$

$$= \sum_{a\in T,F} P(J=T\mid A=a)\quad \tau_2(A=a)$$

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J=T \mid A=a) P(M=m \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e)$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e) \left[\sum_{m\in T,F} P(M=m \mid A=a) \right]$$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} P(J=T \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e) \quad 1$$

$$= \sum_{a\in T,F} \sum_{b\in T,F} P(J=T \mid A=a) P(B=b) \left[\sum_{e\in T,F} P(A=a \mid B=b, E=e) P(E=e) \right]$$

$$= \sum_{a\in T,F} \sum_{b\in T,F} P(J=T \mid A=a) P(B=b)\, \tau_1(A=a, B=b)$$

$$= \sum_{a\in T,F} P(J=T \mid A=a) \left[\sum_{e\in T,F} P(B=b)\, \tau_1(A=a, B=b) \right]$$

$$= \sum_{a\in T,F} P(J=T \mid A=a) \quad \tau_2(A=a) \quad = \boxed{P(J=T)}$$

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b\in T,F} \sum_{e\in T,F} \sum_{a\in T,F} \sum_{m\in T,F} P(J=T \mid A=a) P(M=m \mid A=a) P(A=a \mid B=b, E=e) P(B=b) P(E=e)$$

$$= \sum_{B\in T,F} \sum_{E\in T,F} \sum_{A\in T,F} \sum_{M\in T,F} f_1(A) f_2(M, A) f_3(A, B, E) f_4(B) f_4(E)$$

Conditional probabilities defining the joint = factors

➡

Variable elimination inference can be cast in terms of operations defined over factors

Factors

- **Factor:** is a function that maps value assignments for a subset of random variables to $\Re$ (reals)
- **The scope of the factor:**
  - a set of variables defining the factor
- **Example:**
  - Assume discrete random variables x (with values a1,a2, a3) and y (with values b1 and b2)
  - Factor:

$$\phi(x, y)$$

a1	b1	0.5
a1	b2	0.2
a2	b1	0.1
a2	b2	0.3
a3	b1	0.2
a3	b2	0.4

  - Scope of the factor:

$$\{x, y\}$$

Factor Product

Variables: A,B,C

$$\phi(A,B,C) = \phi(B,C) \circ \phi(A,B)$$

$\phi(B,C)$

b1	c1	0.1
b1	c2	0.6
b2	c1	0.3
b2	c2	0.4

$\phi(A,B)$

a1	b1	0.5
a1	b2	0.2
a2	b1	0.1
a2	b2	0.3
a3	b1	0.2
a3	b2	0.4

$\phi(A,B,C)$

a1	b1	c1	0.5*0.1
a1	b1	c2	0.5*0.6
a1	b2	c1	0.2*0.3
a1	b2	c2	0.2*0.4
a2	b1	c1	0.1*0.1
a2	b1	c2	0.1*0.6
a2	b2	c1	0.3*0.3
a2	b2	c2	0.3*0.4
a3	b1	c1	0.2*0.1
a3	b1	c2	0.2*0.6
a3	b2	c1	0.4*0.3
a3	b2	c2	0.4*0.4

Factor Marginalization

Variables: A,B,C

$$\phi(A,C) = \sum_{B} \phi(A,B,C)$$

a1	b1	c1	0.2
a1	b1	c2	0.35
a1	b2	c1	0.4
a1	b2	c2	0.15
a2	b1	c1	0.5
a2	b1	c2	0.1
a2	b2	c1	0.3
a2	b2	c2	0.2
a3	b1	c1	0.25
a3	b1	c2	0.45
a3	b2	c1	0.15
a3	b2	c2	0.25

a1	c1	0.2+0.4=0.6
a1	c2	0.35+0.15=0.5
a2	c1	0.8
a2	c2	0.3
a3	c1	0.4
a3	c2	0.7

Factor division

A=1	B=1	0.5
A=1	B=2	0.4
A=2	B=1	0.8
A=2	B=2	0.2
A=3	B=1	0.6
A=3	B=2	0.5

A=1	0.4
A=2	0.4
A=3	0.5

A=1	B=1	0.5/0.4=1.25
A=1	B=2	0.4/0.4=1.0
A=2	B=1	0.8/0.4=2.0
A=2	B=2	0.2/0.4=2.0
A=3	B=1	0.6/0.5=1.2
A=3	B=2	0.5/0.5=1.0

Inverse of a factor product

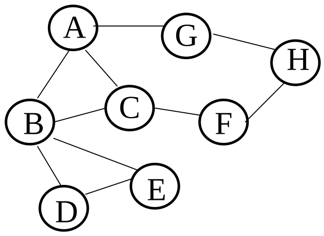
Markov random fields

An undirected network (also called independence graph)

- **Probabilistic models with symmetric dependences**
- G = (S, E)
  - S set of random variables
  - Undirected edges E that define dependences between pairs of variables

Example:

variables A,B ..H



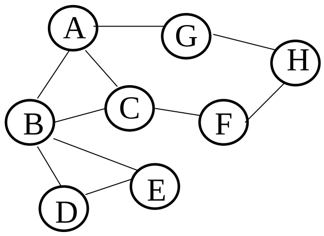
Markov random fields

The full joint of the MRF is defined

$$P(\mathbf{x}) \propto \prod_{c \in cl(x)} \phi_c(\mathbf{x}_c)$$

$\phi_c(x_c)$ - A potential function (defined over variables in cliques/factors)

Example:



Full joint:

$$P(A,B,\ldots H) \sim \phi_1(A,B,C)\phi_2(B,D,E)\phi_3(A,G)\phi_4(C,F)\phi_5(G,H)\phi_6(F,H)$$

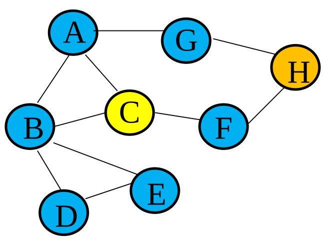
$\phi_c(x_c)$ - A potential function (defined over a clique of the graph)

Markov random fields: independence relations

- **Pairwise Markov property**
  - Two nodes in the network that are not directly connected can be made independent given all other nodes
- **Local Markov property**
  - A set of nodes (variables) can be made independent from the rest of nodes variables given its immediate neighbors
- **Global Markov property**
  - A vertex set A is independent of the vertex set B (A and B are disjoint) given set C if all chains in between elements in A and B intersect C

Markov random fields: independence relations

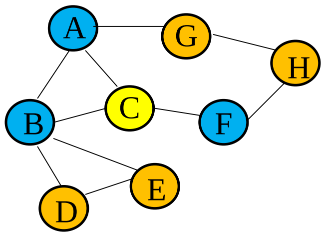
- **Pairwise Markov property**
  - Two nodes in the network that are not directly connected can be made independent given all other nodes
- **Example:**



C and H are independent given the rest of the nodes

Markov random fields: independence relations

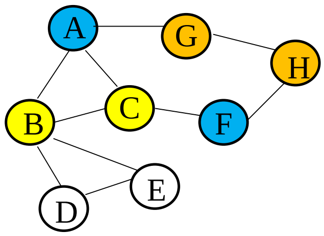
- **Local Markov property**
  – A set of nodes (variables) can be made independent from the rest of nodes variables given its immediate neighbors
- **Example:**



C is independent of {G,H,D,E} given the neighbors of C that is, variables {A,B,F}

Markov random fields: independence relations

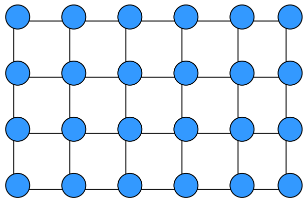
- **Global Markov property**
  – A vertex set A is independent of the vertex set B (A and B are disjoint) given set C if all chains in between elements in A and B intersect C
- **Example:**



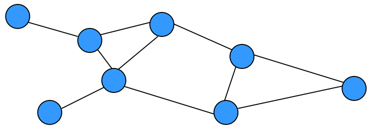
A set {B, C} is independent of {G,H} given the set{A,F}

Markov random fields

- **regular lattice (Ising model)**

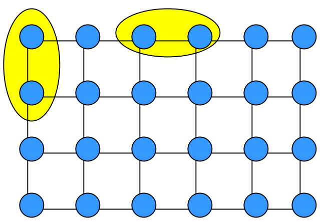


- **Arbitrary graph**

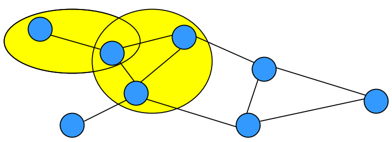


Markov random fields

- **regular lattice (Ising model)**



- **Arbitrary graph**



Markov random fields

- **Joint probability**

$$P(x) \approx \prod_{c \in cl(x)} \phi_c(x_c)$$

$\phi_c(x_c)$ - A potential function (defined over cliques/factors)

- **Typical condition on potential functions:**
  - If $\phi_c(x_c)$ is strictly positive we can rewrite the definition in terms of a log-linear model :

$$P(x) = \frac{1}{Z} \exp\left(- \sum_{c \in cl(x)} E_c(x_c) \right)$$

Energy function

- Gibbs (Boltzman) distribution

$$Z = \sum_{x \in \{x\}} \exp\left(- \sum_{c \in cl(x)} E_c(x_c) \right)$$

- A partition function

Types of Markov random fields

- **MRFs with discrete random variables**
  - Clique potentials can be defined by mapping all clique-variable instances to R (= factors)
  - Example: Assume two binary variables A,B with values {a1,a2,a3} and {b1,b2} are in the same clique c. Then:

$\phi_c(A, B) \cong$

a1	b1	0.5
a1	b2	0.2
a2	b1	0.1
a2	b2	0.3
a3	b1	0.2
a3	b2	0.4

Types of Markov random fields

- **Gaussian Markov Random Field**

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

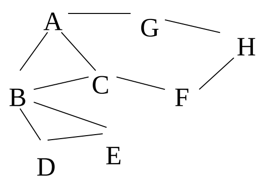
$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

- **Precision matrix** $\boldsymbol{\Sigma}^{-1}$
- **Variables in x are connected in the network only if they have <u>a nonzero entry</u> in the precision matrix**
  - All zero entries are not directly connected

MRF variable elimination inference

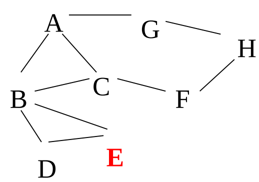
Example:

$$P(B) = \sum_{A,C,D,..H} P(A, B, \ldots H)$$



$$= \frac{1}{Z} \sum_{A,C,D,..H} \phi_1(A,B,C)\phi_2(B,D,E)\phi_3(A,G)\phi_4(C,F)\phi_5(G,H)\phi_6(F,H)$$

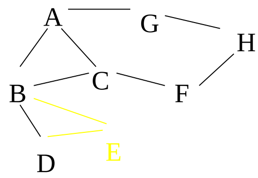
Eliminate E



$$= \frac{1}{Z} \sum_{A,C,D,F,G,H} \phi_1(A,B,C) \underbrace{\left[\sum_E \phi_2(B,D,E) \right]}_{\tau_1(B,D)} \phi_3(A,G)\phi_4(C,F)\phi_5(G,H)\phi_6(F,H$$

MRF variable elimination inference

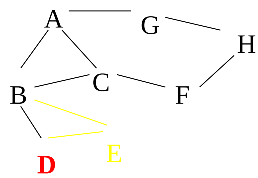
Example (cont):



$$P(B) = \sum_{A,C,D,..H} P(A,B,...H)$$

$$= \frac{1}{Z} \sum_{A,C,D,F,G,H} \phi_1(A,B,C)\tau_1(B,D)\phi_3(A,G)\phi_4(C,F)\phi_5(G,H)\phi_6(F,H)$$

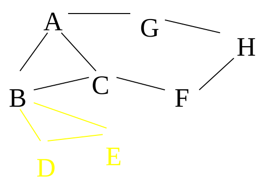
Eliminate D



$$= \frac{1}{Z} \sum_{A,C,F,G,H} \phi_1(A,B,C)\underbrace{\left[\sum_{D} \tau_1(B,D)\right]}_{\tau_2(B)}\phi_3(A,G)\phi_4(C,F)\phi_5(G,H)\phi_6(F,H)$$

MRF variable elimination inference

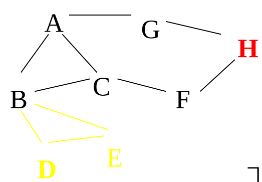
Example (cont):



$$P(B) = \sum_{A,C,D,..H} P(A,B,...H)$$

$$= \frac{1}{Z} \sum_{A,C,F,G,H} \phi_1(A,B,C)\tau_2(B)\phi_3(A,G)\phi_4(C,F)\phi_5(G,H)\phi_6(F,H)$$

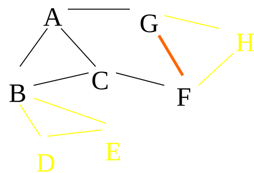
Eliminate H



$$= \frac{1}{Z} \sum_{A,C,F,G} \phi_1(A,B,C)\tau_2(B)\phi_3(A,G)\phi_4(C,F)\underbrace{\left[\sum_{H} \underbrace{\phi_5(G,H)\phi_6(F,H)}_{\tau_3(F,G,H)}\right]}_{\tau_4(F,G)}$$

MRF variable elimination inference

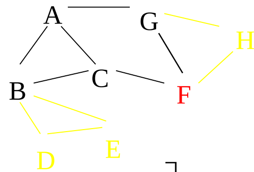
Example (cont):



$$P(B) = \sum_{A,C,D,..H} P(A,B,...H)$$

$$= \frac{1}{Z} \sum_{..,C,F,G} \phi_1(A,B,C)\tau_2(B)\phi_3(A,G)\phi_4(C,F)\tau_4(F,G)$$

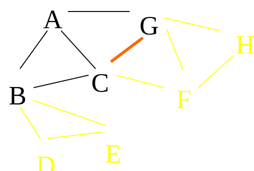
Eliminate F



$$= \frac{1}{Z} \sum_{A,C,G} \phi_1(A,B,C)\tau_2(B)\phi_3(A,G)\underbrace{\left[\underbrace{\sum_F \phi_4(C,F)\tau_4(F,G)}_{\tau_5(C,F,G)}\right]}_{\tau_6(G,C)}$$

MRF variable elimination inference

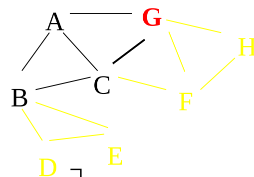
Example (cont):



$$P(B) = \sum_{A,C,D,..H} P(A,B,...H)$$

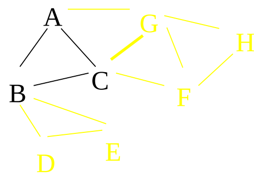
$$= \frac{1}{Z} \phi_1(A,B,C)\tau_2(B)\phi_3(A,G)\tau_6(C,G)$$

Eliminate G



$$= \frac{1}{Z} \sum_{A,C} \phi_1(A,B,C)\tau_2(B)\underbrace{\left[\underbrace{\sum_F \phi_3(A,G)\tau_6(C,G)}_{\tau_7(A,C,G)}\right]}_{\tau_8(A,C)}$$

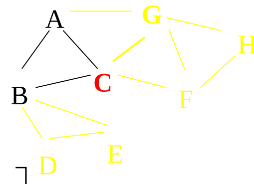
MRF variable elimination inference



Example (cont):

$$P(B) = \sum_{A,C,D,..H} P(A,B,...H)$$

$$= \frac{1}{Z} \sum_{A,C} \phi_1(A,B,C)\tau_2(B)\tau_8(A,C)$$

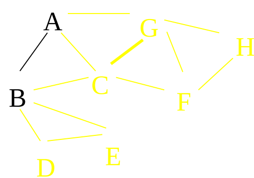


Eliminate C

$$= \frac{1}{Z} \sum_{A} \tau_2(B)\left[\sum_{C} \underbrace{\phi_1(A,B,C)\tau_8(A,C)}_{\tau_9(A,B,C)}\right]$$

$$\underbrace{\left[\sum_{C} \tau_9(A,B,C)\right]}_{\tau_{10}(A,B)}$$

MRF variable elimination inference

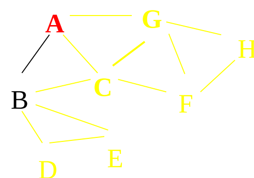


Example (cont):

$$P(B) = \sum_{A,C,D,..H} P(A,B,...H)$$

$$= \frac{1}{Z} \tau_2(B)\tau_{10}(A,B)$$

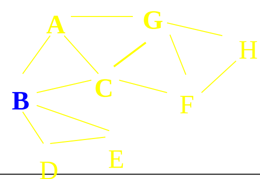
$$= \frac{1}{Z} \tau_2(B)\sum_{A} \tau_{10}(A,B)$$



Eliminate A

$$= \frac{1}{Z} \tau_2(B)\underbrace{\sum_{A} \tau_{10}(A,B)}_{\tau_{11}(B)}$$

$$= \frac{1}{Z} \tau_2(B)\tau_{11}(B)$$



Variable elimination

Depends on the order of variables to eliminate

Question: can we optimize the structures ahead of times so that we can make inferences efficiently and without worrying about the specific variable order?

- **Structures that support efficient inferences: Chains, and trees**