

Markov Networks: Representation

Vibhav Gogate



THE UNIVERSITY OF TEXAS AT DALLAS

Erik Jonsson School of Engineering and Computer Science

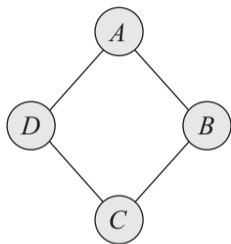
Markov networks: Undirected Graphical models

Model the following distribution using a Bayesian network

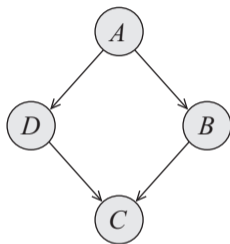
- ▶ $I(A, \{B, D\}, C)$
- ▶ $I(B, \{A, C\}, D)$

Here, we are trying to model how both good and bad information spreads between four mutual friends: Anna (A), Bob (B), Charles (C) and Debbie (D). Anna and Charles used to date each other and the relationship ended badly. Bob and Debbie were married. Unfortunately, they got divorced recently and are not in speaking terms.

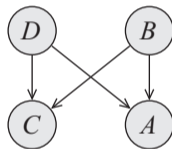
Attempted Bayesian networks



(a)



(b)



(c)

(b) says that A is independent of C given $\{B, D\}$. However, it also implies that B and D are independent given only A but dependent given both A and C .

Markov networks: Undirected Graphical models

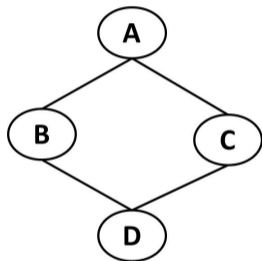
Definition:

- ▶ An undirected graph G
- ▶ A set of factors or functions, denoted by ϕ associated with maximal cliques of G

Example of a factor $\phi(A, B)$

a	b	30
a	$\neg b$	5
$\neg a$	b	1
$\neg a$	$\neg b$	10

Distribution represented by a Markov network



- ▶ Normalized product of all factors (called the **Gibbs distribution**).
- ▶ $\Pr(a, b, c, d) = \frac{1}{Z} \phi(a, b) \times \phi(a, c) \times \phi(b, d) \times \phi(c, d)$
- ▶ Z is a normalizing constant, often called the partition function
- ▶ $Z = \sum_{a,b,c,d} \phi(a, b) \times \phi(a, c) \times \phi(b, d) \times \phi(c, d)$

What does a factor represent?

- ▶ They are not conditional probability distributions. Why? The interactions are not directed.
- ▶ The factors represent affinity between related variables. Higher values are more probable and they trade-off with each other using the partition function Z .
- ▶ Alice and Bob tend to agree with each other.

a	b	30
a	$\neg b$	5
$\neg a$	b	1
$\neg a$	$\neg b$	10

What does a factor represent?

- ▶ They are not conditional probability distributions. Why? The interactions are not directed.
- ▶ The factors represent affinity between related variables. Higher values are more probable and they trade-off with each other using the partition function Z .
- ▶ Alice and Bob tend to agree with each other.

a	b	30
a	$\neg b$	5
$\neg a$	b	1
$\neg a$	$\neg b$	10

What does a factor represent?

- ▶ They are not conditional probability distributions. Why? The interactions are not directed.
- ▶ The factors represent affinity between related variables. Higher values are more probable and they trade-off with each other using the partition function Z .
- ▶ Alice and Bob tend to agree with each other.

a	b	30
a	$\neg b$	5
$\neg a$	b	1
$\neg a$	$\neg b$	10

What does a factor represent?

- ▶ They are not conditional probability distributions. Why? The interactions are not directed.
- ▶ The factors represent affinity between related variables. Higher values are more probable and they trade-off with each other using the partition function Z .
- ▶ Alice and Bob tend to agree with each other.

a	b	30
a	$\neg b$	5
$\neg a$	b	1
$\neg a$	$\neg b$	10

Distribution represented by a Markov network

$\phi(A, B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi(B, C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

$\phi(C, D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi(D, A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

Assignment				UnNormalized	Normalized
a^0	b^0	c^0	d^0	300K	0.04
a^0	b^0	c^0	d^1	300K	0.04
a^0	b^0	c^1	d^0	300K	0.04
a^0	b^0	c^1	d^1	30	4.1E-06
a^0	b^1	c^0	d^0	500	6.9E-05
a^0	b^1	c^0	d^1	500	6.9E-05
a^0	b^1	c^1	d^0	5M	0.69
a^0	b^1	c^1	d^1	500	6.9E-05
a^1	b^0	c^0	d^0	100	1.4E-05
a^1	b^0	c^0	d^1	1M	0.14
a^1	b^0	c^1	d^0	100	1.4E-05
a^1	b^0	c^1	d^1	100	1.4E-05
a^1	b^1	c^0	d^0	10	1.4E-05
a^1	b^1	c^0	d^1	100K	0.014
a^1	b^1	c^1	d^0	100K	0.014
a^1	b^1	c^1	d^1	100K	0.014

The Partition Function vs Local Factors

Compare:

Marginal over A, B is:

a^0	b^0	.13
a^0	b^1	.69
a^1	b^0	.14
a^1	b^1	.04

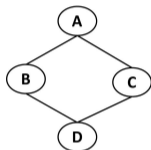
$\phi(A, B)$

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

- ▶ The most likely tuple in $\phi(A, B)$ is the first one. The most likely tuple in the marginal is the second one.
- ▶ **Strength** of other factors is much stronger than that of $\phi(A, B)$ and therefore the influence of the latter is overwhelmed.

Independencies represented by the Markov network

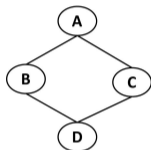
- ▶ **Graph separation:** $I(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if \mathbf{X} and \mathbf{Y} are disconnected after removing \mathbf{Z}



- ▶ What are the independencies represented by this network?
 - ▶ $I(A, \{B, C\}, D)$
 - ▶ $I(B, \{A, D\}, C)$
 - ▶ Compare with the Bayesian network for the example problem.

Independencies represented by the Markov network

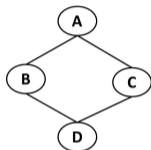
- ▶ **Graph separation:** $I(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if \mathbf{X} and \mathbf{Y} are disconnected after removing \mathbf{Z}



- ▶ What are the independencies represented by this network?
- ▶ $I(A, \{B, C\}, D)$
- ▶ $I(B, \{A, D\}, C)$
- ▶ Compare with the Bayesian network for the example problem.

Independencies represented by the Markov network

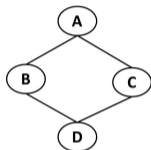
- ▶ **Graph separation:** $I(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if \mathbf{X} and \mathbf{Y} are disconnected after removing \mathbf{Z}



- ▶ What are the independencies represented by this network?
- ▶ $I(A, \{B, C\}, D)$
- ▶ $I(B, \{A, D\}, C)$
- ▶ Compare with the Bayesian network for the example problem.

Independencies represented by the Markov network

- ▶ **Graph separation:** $I(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ only if \mathbf{X} and \mathbf{Y} are disconnected after removing \mathbf{Z}



- ▶ What are the independencies represented by this network?
- ▶ $I(A, \{B, C\}, D)$
- ▶ $I(B, \{A, D\}, C)$
- ▶ Compare with the Bayesian network for the example problem.

Independencies represented by the Markov network

- ▶ If \Pr is a Gibbs distribution of an undirected graph G , then G is an I-map of \Pr .
- ▶ If G is an I-map of a positive distribution \Pr , then \Pr is a Gibbs distribution.

Theorem (Hammersley-Clifford Theorem)

Let \Pr be a positive distribution, then \Pr is a Gibbs distribution of a Markov network G if and only if G is an I-map of \Pr .

Independencies represented by the Markov network

- ▶ Remember Bayesian networks. We had local independencies (Markov) and global independencies (d-separation).
- ▶ Analogously, we can define local independencies via Markov blanket (neighbors of node).

$$I(X, \text{Neighbors}(X), \text{Non_Neighbors}(X))$$

- ▶ Global independencies via Graph separation follow from these local independencies
- ▶ Building minimal I-maps
 1. For each node X find a minimal set of nodes $MB(X)$ such that the local independence property is satisfied.
 2. Connect X with $MB(X)$

Bayesian vs Markov Networks

Property	Bayesian networks	Markov networks
Functions	CPTs	Factors/Potentials
Distribution	Product of CPTs	normalized product
Cycles	Not Allowed	Allowed
Partition function	$Z = 1$	$Z = ?$
Independencies Check	D-separation	Graph-separation
Compact models	Class-BN	Class-MN

- ▶ Class-BN \neq Class-MN

Converting Bayesian to Markov networks

- ▶ **Moralization:** Connect all parents of a node to each other. Remove directionality.
- ▶ Moralized graph MG of a Bayesian network G is a **minimal** I-map of the distribution represented by G .
- ▶ Moralization results in loss of some independence information.
- ▶ Example: $X \rightarrow Z \leftarrow Y$.
- ▶ Markov network? Clique over X, Y and Z
- ▶ Loss: $I(X, \emptyset, Y)$ no longer holds in the Markov network.

Converting Bayesian to Markov networks

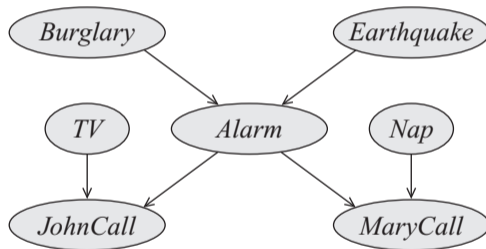
- ▶ **Moralization:** Connect all parents of a node to each other. Remove directionality.
- ▶ Moralized graph MG of a Bayesian network G is a **minimal** I-map of the distribution represented by G .
- ▶ Moralization results in loss of some independence information.
- ▶ Example: $X \rightarrow Z \leftarrow Y$.
- ▶ Markov network? Clique over X, Y and Z
- ▶ Loss: $I(X, \emptyset, Y)$ no longer holds in the Markov network.

Converting Bayesian to Markov networks

- ▶ **Moralization:** Connect all parents of a node to each other. Remove directionality.
- ▶ Moralized graph MG of a Bayesian network G is a **minimal** I-map of the distribution represented by G .
- ▶ Moralization results in loss of some independence information.
- ▶ Example: $X \rightarrow Z \leftarrow Y$.
- ▶ Markov network? Clique over X , Y and Z
- ▶ Loss: $I(X, \emptyset, Y)$ no longer holds in the Markov network.

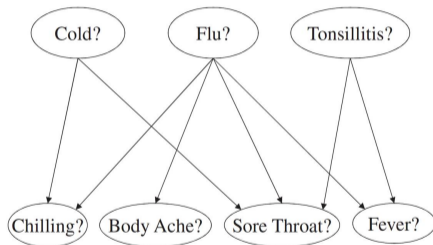
Converting Bayesian to Markov networks

Draw an equivalent Markov network.



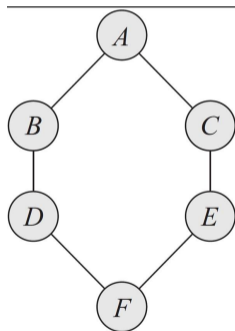
Converting Bayesian to Markov networks

Draw an equivalent Markov network.



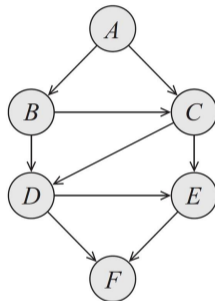
Converting Markov to Bayesian networks

- ▶ Non-trivial and hard!
- ▶ Requires triangulating the Markov network.
- ▶ Example: Convert the following to a Bayesian network



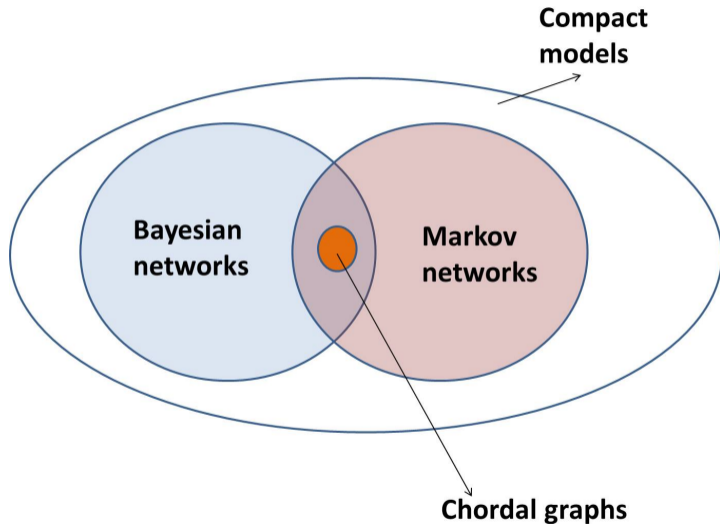
Converting Markov to Bayesian networks

Equivalent Bayesian network that is a minimal I-map of the Markov network.



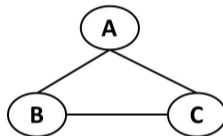
- ▶ Loss of independence assumptions: $I(C, \{A, F\}, D)$ does not hold in the Bayesian network.

Classes of Compact models



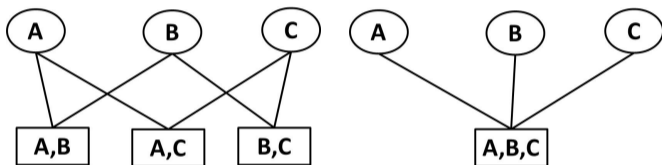
Factor Graphs

- ▶ A Markov network does not reveal its Gibbs distribution. We need to know the parameters! Example:



- ▶ Are the factors $\phi(A, B)$, $\phi(A, C)$, $\phi(B, C)$ or is there a single factor $\phi(A, B, C)$?
- ▶ Factor graph is a fine grained factorization. It is a bi-partite graph.

Factor Graphs



Quiz: How to convert a Markov network to a factor graph?

Log-linear models

- ▶ Convert each tuple in a function to a formula.
- ▶ If the formula is True then weight equals log of the function value and 0 otherwise.

$$\Pr(\mathbf{x}) = \frac{1}{Z} \exp \left(\sum_{f_i} w_i \times I_{\mathbf{x}}(f_i) \right)$$

$I_{\mathbf{x}}(f_i) = 1$ if \mathbf{x} satisfies f_i and 0 otherwise.

Var 1	Var 2	ϕ	Formula	Weight
A	B	30	$A \wedge B$	$\log(30)$
A	$\neg B$	5	$A \wedge \neg B$	$\log(5)$
$\neg A$	B	1	$\neg A \wedge B$	$\log(1)$
$\neg A$	$\neg B$	30	$\neg A \wedge \neg B$	$\log(30)$

Log-linear models: Mapping logic to Probabilities

If two formulas have the same weight, we can merge the two:

Var 1	Var 2	ϕ	Formula	Weight
A	B	30	$A \wedge B$	$\log(30)$
A	$\neg B$	5	$A \wedge \neg B$	$\log(5)$
$\neg A$	B	1	$\neg A \wedge B$	$\log(1)$
$\neg A$	$\neg B$	30	$\neg A \wedge \neg B$	$\log(30)$

Merging first and fourth formula, we have a new $[formula, weight]$ pair:
 $[(A \wedge B) \vee (\neg A \wedge \neg B), \log(30)]$.

Significant compaction can be achieved in some cases.

Building Graphical models by hand: Skills

- ▶ Identify relevant variables and their possible values.
- ▶ Build the network structure (a Bayesian network or a factor graph)
[Issues: CPTs/Factors can be large]
- ▶ Define the specific numbers in CPTs or factors
[Issues: How significant are the numbers? Will a 0.1 instead of 0.01 matter?]
- ▶ Identify the specific queries that needs to be answered (application dependent)

Types of Queries

- ▶ Probability of evidence
- ▶ Prior vs Posterior marginals
- ▶ Most probable explanation
- ▶ Maximum a posteriori hypothesis (MAP)

