

# From Bayesian Networks to Markov Networks

Sargur Srihari

[srihari@cedar.buffalo.edu](mailto:srihari@cedar.buffalo.edu)

# Topics

- Bayesian Networks and Markov Networks
- From BN to MN: Moralized graphs
- From MN to BN: Chordal graphs

# Bayesian Networks and Markov Networks

- Bayesian networks and Markov networks are languages for representing independencies
- Each can represent independence constraints that other cannot
  - E.g., in misconception example, some independences in MN cannot be represented in a BN
- Next: insight into relationship between the two representations
  - Begin by how a distribution in one framework can be represented in the other

# From Bayesian Networks to MNs

- Two perspectives

1. Given a BN  $\mathcal{B}$ , ie., a graph with CPDs, how to represent the distribution  $P_{\mathcal{B}}$  as a parameterized MN

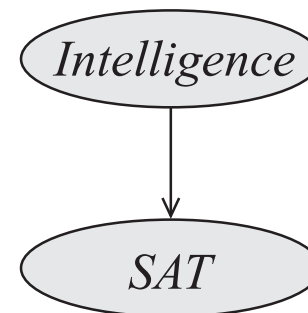
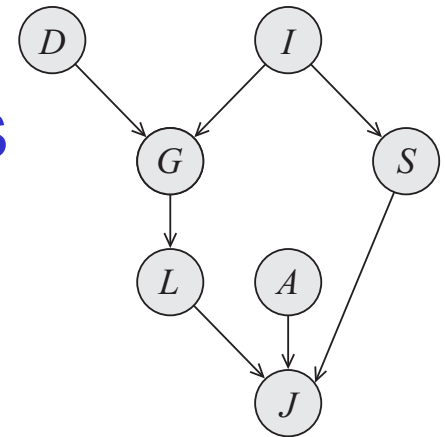
OR

2. Given graph  $\mathcal{G}$  how to represent independencies in  $\mathcal{G}$  using an undirected graph  $\mathcal{H}$

# BN is Gibbs with $Z=1$

- Consider a distribution  $P_{\mathcal{B}}$  where  $\mathcal{B}$  is parameterized BN over graph  $\mathcal{G}$ 
  - Parameters of  $\mathcal{B}$  can be viewed as parameters for a Gibbs distribution
  - Take each CPD  $p(X_i | Pa_{X_i})$  and view it as a factor of scope  $\{X_i, Pa_{X_i}\}$
  - Its partition function is 1
    - Since it is already normalized

$$Z = \sum_{\xi} \prod_i p(X_i | Pa_{X_i})$$



$I$	$S$	$P(I, S)$
$i^0$	$s^0$	0.665
$i^0$	$s^1$	0.035
$i^1$	$s^0$	0.06
$i^1$	$s^1$	0.24.

- More importantly, a BN conditioned on evidence  $E=e$  also induces a Gibbs distribution: 5

# BN with evidence $e$ is Gibbs with $Z=P(e)$

- Consider a BN defined by original factors reduced to context  $E=e$
- $B$  is a BN over  $\chi$  and  $E=e$  an observation. Let  $W=\chi-E$ .
  - Then  $P_B(W|e)$  is a Gibbs distribution with factors  $\Phi=\{\phi_{X_i}\}$   $X_i \in \chi$  where  $\phi_{X_i}=P_B(X_i|Pa_{X_i})[E=e]$
  - Partition function for Gibbs distribution is  $P(e)$ . Proof follows:

$$P_B(\chi) = \prod_{i=1}^N P_B(X_i | Pa_{X_i})$$

$$P_B(W | E = e) = \frac{P_B(W)[E = e]}{P_B(E = e)} = \frac{\prod_{i=1}^N P_B(X_i | Pa_{X_i})[E = e]}{\sum_W P_B(\chi)[E = e]} = \frac{\prod_{i=1}^N P_B(X_i | Pa_{X_i})[E = e]}{\sum_W \prod_{i=1}^N P_B(X_i | Pa_{X_i})[E = e]}$$

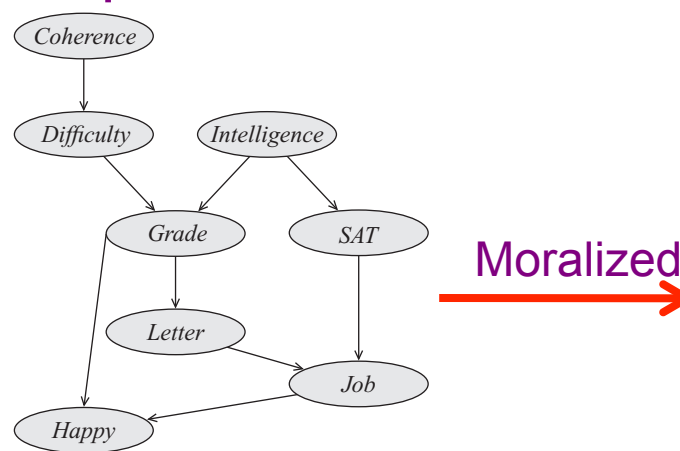
- Thus any BN conditioned on evidence can be regarded as a Markov network
  - and use techniques developed for MN analysis

# Structure of I-map for a BN

- In the construction
  - Each CPD  $p(X_i | Pa_{X_i})$  of the BN is viewed as a factor of scope  $\{X_i, Pa_{X_i}\}$
  - We have created a factor for each family of  $X_i$  containing all the variables in the family
- Thus in undirected I-map we need to have an edge between
  1.  $X_i$  and each of its parents
  2. Between all parents of  $X_i$

# Moralized Graph

- Moral graph  $M(\mathcal{G})$  of a Bayesian network  $\mathcal{G}$  is an undirected graph over  $\mathcal{X}$ 
  - Contains an undirected edge between  $X$  and  $Y$  if
    - a) There is a directed edge between them in the BN OR
    - b)  $X$  and  $Y$  are both parents of the same node



(a)  $M(\mathcal{G})$  has extra links

- It follows that

1.  $M(\mathcal{G})$  is a minimal I-map for  $\mathcal{G}$
2. If  $\mathcal{G}$  is moral then  $M(\mathcal{G})$  is a perfect map of  $\mathcal{G}$



# Parameterization

- No direct way of converting parameters of a BN to parameters of the moralized MN
- Involves setting up of a likelihood function and gradient descent
- Computationally hard due to the need for computing the partition function at each step
- Discussed further in “Learning Undirected Models”

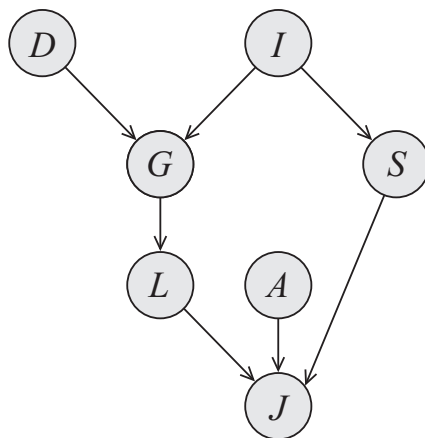
# Soundness of D-separation

- Connection of BNs and MNs provides us with tools for proving soundness of D-separation in BNs
  - i.e., D-separation in BN implies separation in moralized graph
  - So that we can Leverage the idea of separation in MNs

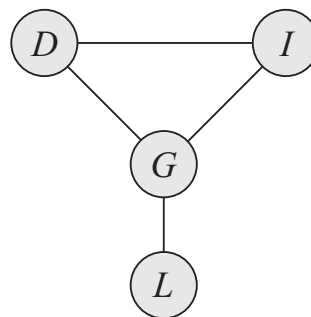
# Soundness of D-separation using MNs

- If  $X, Y, Z$  are disjoint nodes in a BN  $\mathcal{G}$
- Let  $U = X \cup Y \cup Z$  and  $\mathcal{G}^+(U)$  is the induced BN over  $U$  and its ancestors
- Let  $\mathcal{H}$  be the moralized graph  $M[\mathcal{G}^+(U)]$
- Then  $\text{d-sep}_{\mathcal{G}}(X; Y|Z)$  if and only if  $\text{sep}_{\mathcal{H}}(X; Y|Z)$ 
  - D-separation in BN implies separation in moralized graph

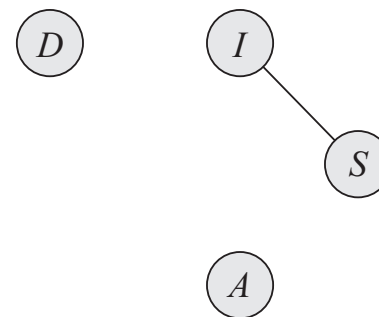
Bayesian Network  $\mathcal{G}$



Markov Network  $M[\mathcal{G}^+[D, I, L]]$



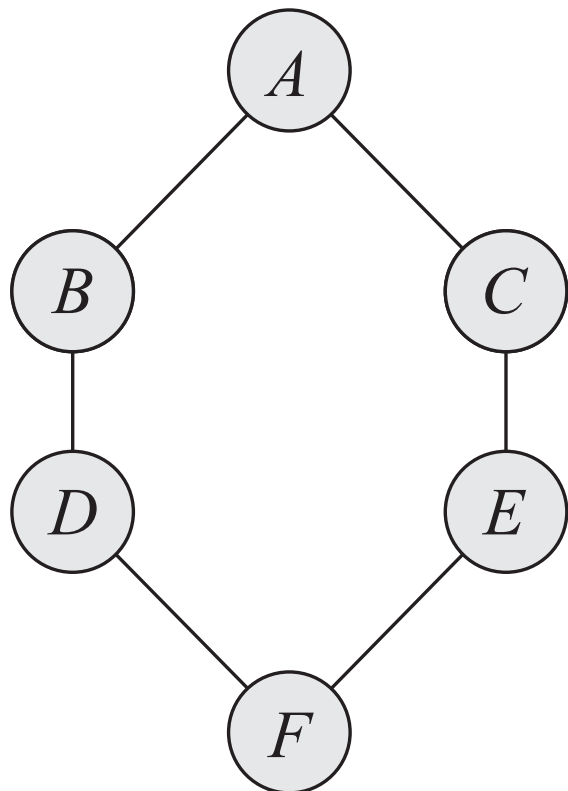
Markov Network  $M[\mathcal{G}^+[D, I, A, S]]$



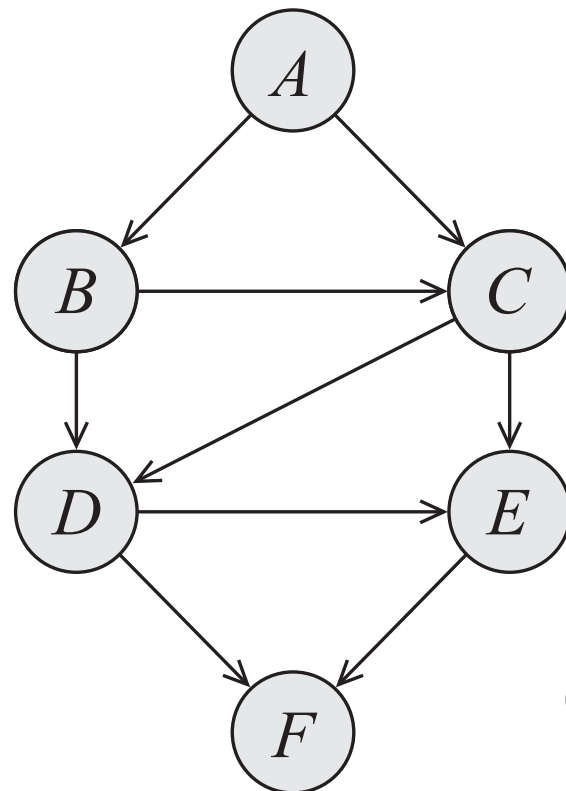
# From MNs to Bayesian Networks

- Finding a BN that is a minimal I-map for a MN
- Significantly more difficult
  - Conceptually, computationally
  - Bayesian network can be considerably larger

# From MN, its minimal I-map BN



MN with a loop  
(a)



Chordal

Minimal I-map BN  
(b)

# Chordal Graphs

- When can a set of independences be perfectly represented by both a Bayesian network and a Markov network?
  - This class is precisely the class of undirected chordal graphs
- Chordal definition
  - Let  $X_1 — X_2 \text{ } \dots \text{ } X_k — X_1$  be a *loop* in the graph
  - A *chord* in the loop is an edge connecting  $X_i$  and  $X_j$  that are non-consecutive
  - An undirected graph is *chordal* if any loop for  $k \geq 4$  has a chord

# Chordal Graph Property

- Let  $\mathcal{H}$  be a chordal MN
- Then there is a BN such that  $I(\mathcal{H})=I(G)$