

# Markov Network Independencies

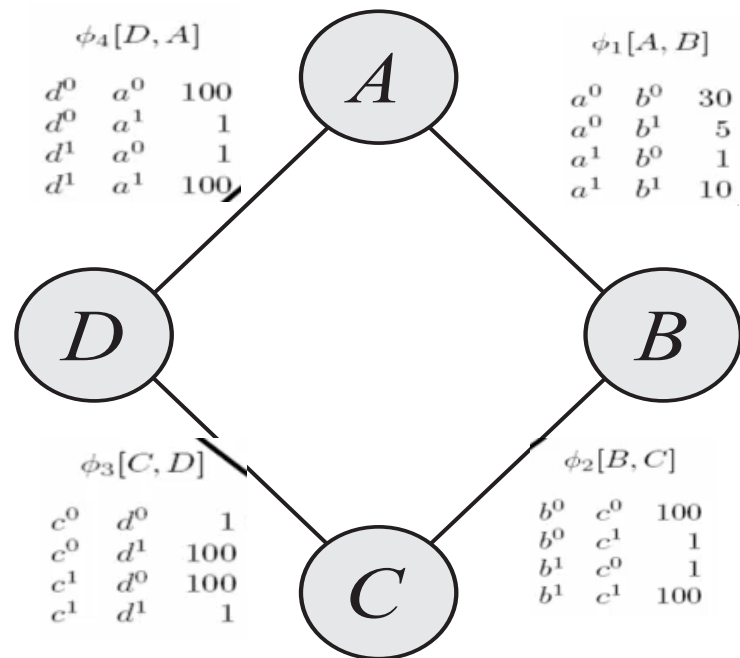
Sargur Srihari  
srihari@cedar.buffalo.edu

# Topics

- Markov Network Independencies
  - Basic Independencies
  - Independencies Revisited
  - From Distributions to Graphs

# Markov network captures independencies of interactions

Misconception MN with factors



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

where

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

As with BNs tight connection between factorization and independence properties:

$P$  supports  $(X \perp Y | Z)$  iff we can write distribution as

$$P(\chi) = \phi_1(X, Z) \phi_2(Y, Z)$$

Consider now:  $P(A,B,C,D) = \left[ \frac{1}{Z} \phi_1(A,B) \cdot \phi_2(B,C) \right] \phi_3(C,D) \cdot \phi_4(D,A)$   
 From this decomposition we can infer that  $(B \perp D | A, C)$

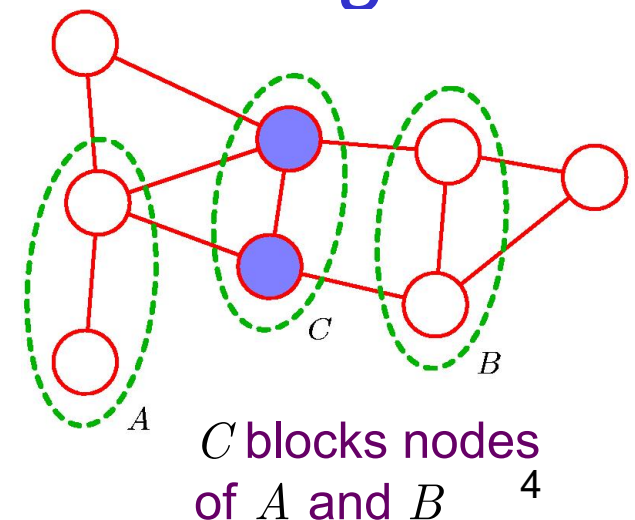
*Proof:* the first grouping yields a factor with  $\{D, \{A, C\}\}$

The second grouping yields a factor with  $\{B, \{A, C\}\}$

Similarly we can infer  $(A \perp C | B, D)$

# Basic Independencies

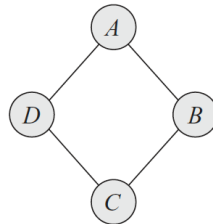
- As in Bayesian Networks, graph structure in a Markov network encodes a set of independence assumptions
- In a MN Probabilistic influence flows along the undirected paths in the graph and “*blocked*” if we condition on intervening nodes
  - i.e., we know their values
- We state this formally, next



# Active Path Definition

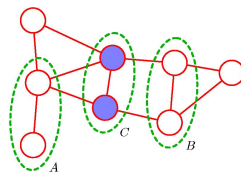
- Let  $\mathcal{H}$  be a Markov network with nodes  $\chi = \{X_1, \dots, X_n\}$ ,
- Let  $X_1 - \dots - X_k$  be a path in  $\mathcal{H}$
- Let  $Z \subseteq \chi$  be a set of observed variables
- A Path  $X_1 - \dots - X_k$  is active given  $Z$  if none of  $X_i$  is in  $Z$

• Ex 1:



If the observed set  $Z = \{B\}$ , path  $A-D-C$  is active

• Ex 2:



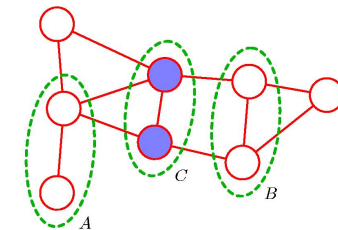
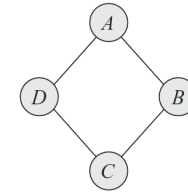
If the observed set  $Z = \{C\}$ , paths between nodes of  $\{A\}$  and  $\{B\}$  are inactive

- We can define separation in the graph when there is no active path, next

# Separation and Global Independencies

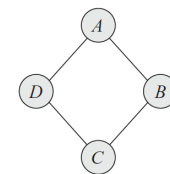
1. Set of nodes  $Z$  separates sets  $X$  and  $Y$  denoted  $\text{sep}_{\mathcal{H}}(X; Y | Z)$  if there is no active path between any  $X \in X$  and  $Y \in Y$

- **Ex 1:**  $Z = \{B, D\}$  separates  $A$  and  $C$ 
  - i.e.,  $\text{sep}_{\mathcal{H}}(A; C | B, D)$  or there is no active path between  $A$  and  $C$
- **Ex 2:**  $C$  separates  $A$  and  $B$



2. Global independencies associated with  $\mathcal{H}$  are

- $I(\mathcal{H}) = \{(X \perp Y | Z) : \text{sep}_{\mathcal{H}}(X; Y | Z)\}$
- Independencies in  $I(\mathcal{H})$  guaranteed to hold for every distribution  $P$  over  $\mathcal{H}$



$$I(\mathcal{H}) = \{(A \perp C | B, D), (B \perp D | A, C)\}$$

# Definition of Separation Leads to a disadvantage

- With a superset of  $Z$ , separation still holds
  - If  $sep_{\mathcal{H}}(X; Y | Z)$  then  $sep_{\mathcal{H}}(X; Y | Z')$  for any  $Z' \supset Z$
- If separation is taken as definition of independencies, we restrict ability to encode non-monotonic independence relations
  - Non-monotonic reasoning is quite useful
    - E.g., intercausal reasoning with BNs
      - Two diseases are independent, but dependent given some common symptom
  - Such independence properties cannot be expressed as a Markov network

# Factorization and Independencies

- Can show connection between independence properties implied by a Markov structure and factorizing a distribution over the graph
- Analogous to Bayesian Networks
  - Let  $\mathcal{G}$  be a BN for a set of random variables  $\chi$  and  $P$  be a distribution over  $\chi$ .
  - If  $P$  factorizes according to  $\mathcal{G}$ , i.e., product of CPDs, then  $\mathcal{G}$  is an I-map of  $P$ 
    - i.e., independencies  $I(\mathcal{G}) \subseteq I$



# Formalizing independencies in MNs and distributions

- Gibbs Distribution

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

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

where

$$\tilde{P}(X_1, \dots, X_n) = \prod_{i=1}^m \phi_i(D_i)$$

is an unnormalized measure and

$$Z = \sum_{X_1, \dots, X_n} \tilde{P}(X_1, \dots, X_n) \text{ is a normalizing constant}$$

called the partition function

$D_i$  are sets of random variables

# Soundness of Separation Criterion

- *Theorem 1* (from factorization to independencies):
  - Let  $P$  be distributed over  $\chi = \{X_1, \dots, X_n\}$  and  $\mathcal{H}$  a Markov structure over  $\chi$
  - If  $P$  is a Gibbs distribution that factorizes over  $\mathcal{H}$ , (i.e., every  $D_i$  in  $\mathcal{H}$  is a clique), then  $\mathcal{H}$  is an I-map for  $P$  (i.e., every independency in  $\mathcal{H}$  holds in  $P$ )
- *Theorem 2: Hammersley-Clifford*  
(other direction: from independencies to factorization)
  - if  $\mathcal{H}$  is an I-map for  $P$  then  $P$  factorizes over  $\mathcal{H}$
  - Holds only for positive distributions ( $P > 0$  for every assignment)

# Positive Distribution

- A distribution  $P$  is said to be positive if for all events  $\alpha \in \mathcal{S}$  such that  $\alpha \neq \emptyset$  we have that  $P(\alpha) > 0$
- Ex: A non-positive distribution
  - 16 possible values
    - Distribution  $P$ : 8 have value  $1/8$  rest are zero

$X_1$	$X_2$	$X_3$	$X_4$	$P(\mathbf{X})$
0	0	0	0	1/8
1	0	0	0	1/8
1	1	0	0	1/8
1	1	1	0	1/8
0	0	0	1	1/8
0	0	1	1	1/8
0	1	1	1	1/8
1	1	1	1	1/8

Rest 8 probs are 0

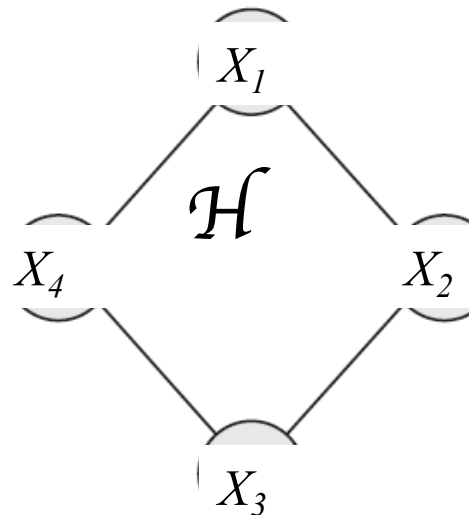
# Non-positive distribution consistent with $\mathcal{H}$

- Four binary random variables
- Global Independencies in graph  $\mathcal{H}$ :

- Consider  $X_1 - X_2 - X_3 - X_4 - X_1$ 
  - implies  $(X_1 \perp X_3 | X_2, X_4)$

- $P$  also satisfies this

- For the assignment  $X_2=1, X_4=0$
- $P(X_1=1 | X_2=1, X_4=0)=1$ 
  - Rest are zero
  - Thus  $X_1$  is independent of  $X_3$

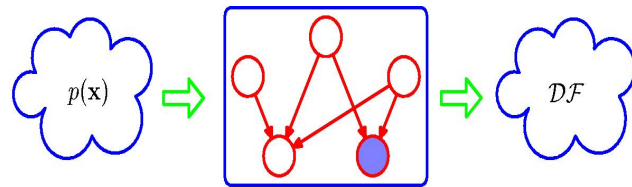


$X_1$	$X_2$	$X_3$	$X_4$	$P(\mathbf{X})$
0	0	0	0	1/8
1	0	0	0	1/8
1	1	0	0	1/8
1	1	1	0	1/8
0	0	0	1	1/8
0	0	1	1	1/8
0	1	1	1	1/8
1	1	1	1	1/8

Rest 8 probs are 0

- Global independencies hold  $\Rightarrow \mathcal{H}$  is an I-map for  $P$
- But  $P$  does not factorize according to  $\mathcal{H}$  (Proof by contradiction)<sup>12</sup>

# Graphical Model as Filter



$p(x)$  is allowed to pass through only if  
It satisfies independencies in graph  
This set is denoted  $DF$  or  $UF$  (for BN or MN)

- UI is set of distributions that are consistent with set of conditional independence statements read from the undirected graph using graph separation
- UF are set of distributions that can be expressed as factorization of the form

$$p(\mathbf{x}) = \frac{1}{Z} \prod_c \psi_c(\mathbf{x}_c)$$

- *Hammersley-Clifford theorem* states that UI and UF are identical

# Independencies in Bayesian Networks

- Bayesian networks have two types of independencies
  - Local independencies
    - Each node is independent of its non-descendants given its parents
  - Global independencies
    - Induced by d-separation
- These two sets of independencies are equivalent
  - One implies the other

# Three Independencies of an MN

## 1. Pairwise Independencies (defined next slide)

- Pairwise  $I_p(\mathcal{H})$

## 2. Local independencies (defined shortly)

- Markov Blanket  $I_\ell(\mathcal{H})$

## 3. Global independency $I(\mathcal{H})$

- Identify three sets of nodes  $A$ ,  $B$  and  $C$
- To test conditional independence property

$$A \perp B \mid C$$

- Consider all possible paths from nodes in set  $A$  to nodes in set  $B$ 
  - If all such paths pass through one or more nodes in  $C_{15}$  then path is blocked and independence holds

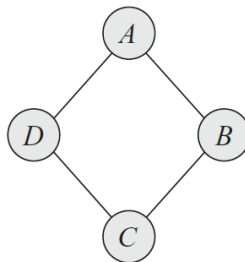
# Pairwise Independence $I_p(\mathcal{H})$

- If  $\mathcal{H}$  is a MN its *pairwise* independencies are

$$I_p(\mathcal{H}) = \{(X \perp Y | \chi - \{X, Y\}) : X - Y \notin \mathcal{H}\}$$

Meaning: When  $X, Y$  are **not** directly connected  
i.e.,  $X - Y \notin \mathcal{H}$ ,  
they are **independent** given all other variables

– Example:

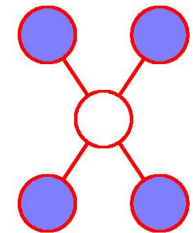


$$I_p(\mathcal{H}) = \{(A \perp C | B, D), (B \perp D | A, C)\}$$



# Markov Blanket Independency $I_{\ell}(\mathcal{H})$

- Analogous to local independencies in Bayesian networks
  - We can block all influences by conditioning on its immediate neighbors
    - Node is conditionally independent of all nodes given its immediate neighbors
- For graph  $\mathcal{H}$  the Markov blanket of  $X$  in  $\mathcal{H}$  is the set of neighbors of  $X$  in  $\mathcal{H}$
- Local independencies associated are
  - $I_{\ell}(\mathcal{H}) = \{(X \perp \chi - \{X\} - MB_{\mathcal{H}}(X) | MB_{\mathcal{H}}(X)) : X \in \chi\}$

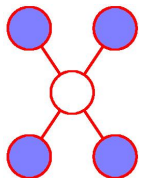
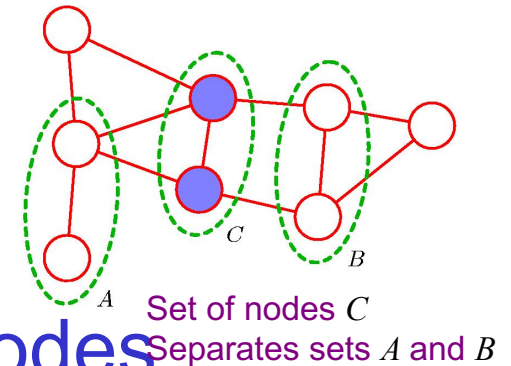


# Relationship between Markov properties

- Three independencies of network structure  $\mathcal{H}$
- $I_p(\mathcal{H})$  is strictly weaker than  $I_\ell(\mathcal{H})$  is strictly weaker than  $I(\mathcal{H})$
- For positive distributions all three are equivalent

# Separation in Markov Networks

- Markov network encodes a set of conditional independencies
- Probabilistic influence flows
  - in undirected paths
- Blocked if we condition on intervening nodes
  - Every path from any node in  $A$  to  $B$  passes through  $C$
  - No explaining away
    - Testing for independence simpler than in directed graphs
  - Alternative view
    - Remove all nodes in set  $C$  together with all their connecting links
    - If no paths from  $A$  to  $B$  then conditional independence holds
- Markov blanket



- A node is conditionally independent of all other nodes conditioned only on its neighbors

# Factorization Properties

- Factorization rule corresponds to conditional independence test
- Notion of locality needed
- Consider two nodes  $x_i$  and  $x_j$  not connected by a link
  - They are conditionally independent given all other nodes in graph
    - Because there is no direct path between them and
    - All other paths pass through nodes that are observed and hence those paths are blocked
  - Expressed as
 

$$p(x_i, x_j \mid \mathbf{x}_{\setminus\{i,j\}}) = p(x_i \mid \mathbf{x}_{\setminus\{i,j\}}) p(x_j \mid \mathbf{x}_{\setminus\{i,j\}})$$
  - Where  $\mathbf{x}_{\setminus\{i,j\}}$  denotes set  $\mathbf{x}$  of all variables with  $x_i$  and  $x_j$  removed
- For conditional independence to hold
  - factorization is such that  $x_i$  and  $x_j$  do not appear in the same factor
    - No path between them other than going through others
  - leads to graph concept of clique

