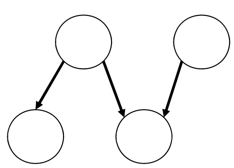
Undirected Graphical Models 3
Bayesian Networks and Markov Networks

1

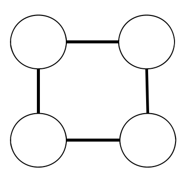
Bayesian Networks and Markov Networks

We have now seen two types of graphical models for representing joint probability distributions:

Bayesian Networks



Markov Networks

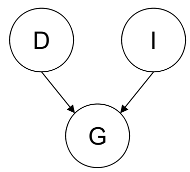


Can we convert from one type to the other?

2

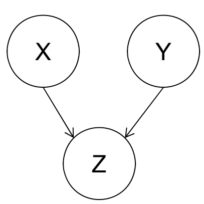
Bayesian Networks and Markov Networks

- Can a Markov network represent any Bayesian network? No!
- Can't use a Markov network to represent a distribution corresponding a Bayesian network v-structure

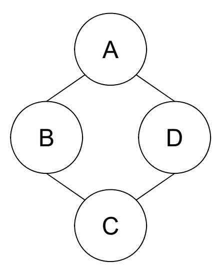


3

Bayesian Networks and Markov Networks



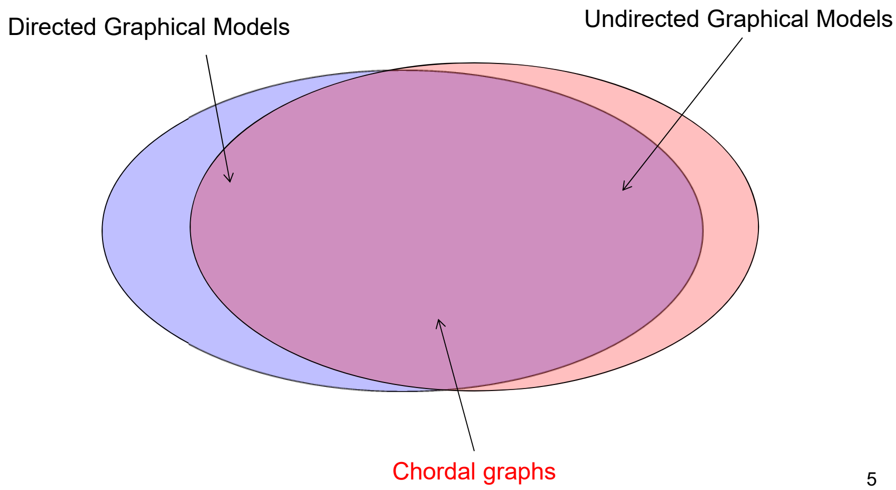
Bayesian networks can represent independence constraints that Markov networks cannot



Markov networks can represent independence constraints that Bayesian networks cannot

4

Bayesian Networks and Markov Networks



5

Bayesian Networks to Markov Networks

6

BNs to MNs

To convert a Bayesian Network B to a Markov Network H, we can think of B as a Gibbs distribution with:

- Each factor ϕ_{X_i} in H corresponds to each conditional probability table $P(X_i|Parents(X_i))$ in B
- The scope of factor ϕ_{X_i} is $X_i \cup Parents(X_i)$
- The partition function $Z = 1$

7

BNs to MNs

How do we create an undirected graph that is an I-map for the distribution P_B represented by the BN?

Need to moralize each factor:

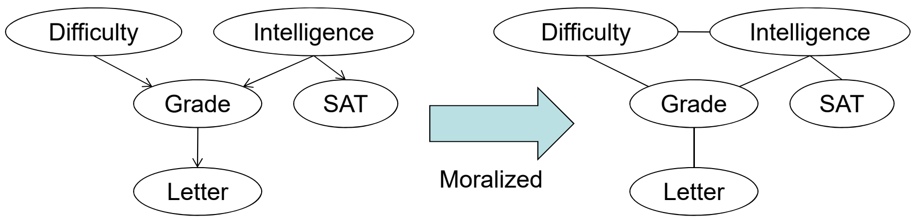
- Add an edge between X_i and its parents
- Add an edge between all the parents of X_i

8

BNs to MNs

The moral graph $M[G]$ of a Bayesian network structure G over $\boldsymbol{X}$ is the undirected graph over $\boldsymbol{X}$ that contains an undirected edge between X and Y if:

- There is a directed edge between them in G (in either direction)
- Or X and Y are both parents of the same node



9

BNs to MNs

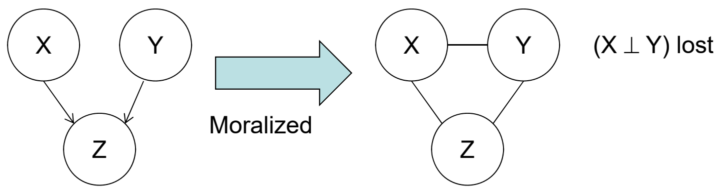
Let G be any Bayesian network graph. The moralized graph $M[G]$ is a minimal I-map for G.

[See proof of Proposition 4.8 in textbook]

10

BNs to MNs

- The addition of the moralizing edges to the MN H leads to the loss of independence information implied by the graph structure

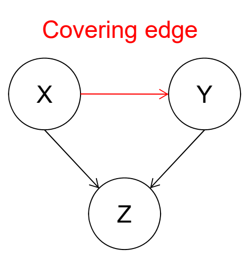


- Only happens if moralization adds new edges into the graph

11

BNs to MNs

A Bayesian network G is moral if it contains no immoralities ie. for any pair of variables X, Y that share a child, there is a covering edge between X and Y



12

BNs to MNs

- If the directed graph G is moral, then its moralized graph $M[G]$ is a perfect map of G.
- Unfortunately, very few directed graphs are moral

13

Markov Networks to Bayesian Networks

14

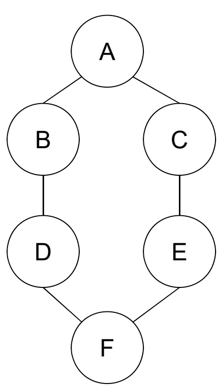
MNs to BNs

- Much harder (conceptually and computationally) to find a Bayesian network that is a minimal I-map for a Markov network
- Resulting Bayesian network might have (many) more edges than the Markov network

15

MNs to BNs

Example: Find a Bayesian network I-map for the given Markov Network



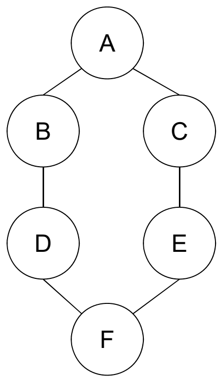
Build-Minimal-I-Map (from Section 3.4.1):
To find a minimal I-map for a distribution *P*:
- Pick a variable ordering
- For each variable X_i in the ordering:
    - Find some minimal subset $\boldsymbol{U}$ of $\{X_1, \dots, X_{i-1}\}$ to be X_i's parents in G such that $\{X_i \perp \{X_1, \dots, X_{i-1}\} - \boldsymbol{U} | \boldsymbol{U}\}$

Note: $(X_i \perp \emptyset | \{X_1, \dots X_{i-1}\})$ is trivially true

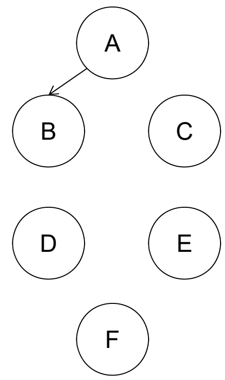
16

MNs to BNs

Example: Find a Bayesian network I-map for the given Markov Network



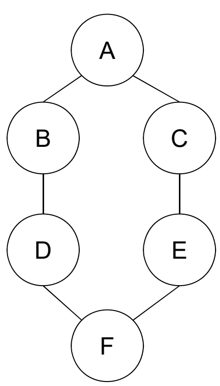
Ordering: A,B,C,D,E,F



17

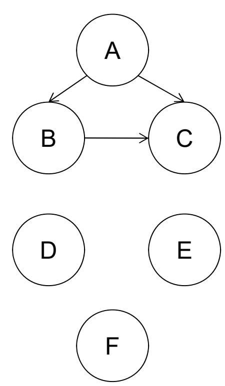
MNs to BNs

Example: Find a Bayesian network I-map for the given Markov Network



Ordering: A,B,C,D,E,F

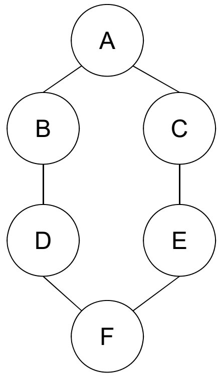
(C not $\perp$ B | A), so must add B as parent of C



18

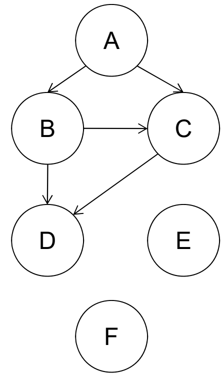
MNs to BNs

Example: Find a Bayesian network I-map for the given Markov Network



Ordering: A,B,C,D,E,F

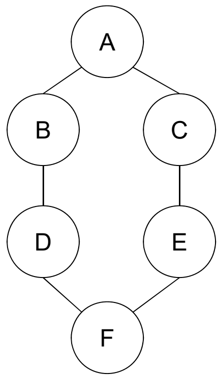
(D not ⊥ C | B), so must Add C as parent of B.
(D ⊥ A | {B,C})



19

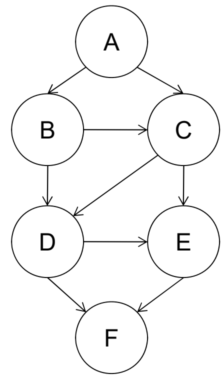
MNs to BNs

Example: Find a Bayesian network I-map for the given Markov Network



Ordering: A,B,C,D,E,F

And so on…
Notice the edges added which result in a set of triangles.
The graph is chordal (defn to follow)



20

MNs to BNs

- Let X_1—X_2—X_3—X_4 be a loop in a graph
- A chord in the loop is an edge connecting X_i and X_j for two nonconsecutive nodes X_i, X_j.
- An undirected graph H is said to be chordal if any loop X_1—X_2—…—X_k—X_1 for $K \geq 4$ has a chord

21

MNs to BNs

Theorem 4.10: Let H be a Markov network structure, and let G be any Bayesian network minimal I-map for H. Then G can have no immoralities. [Proof omitted]

22

MNs to BNs

Any nontriangulated loop of length at least 4 in a Bayesian network graph necessarily contains an immorality.

Corollary 4.3: Let H be a Markov network structure, and let G be any minimal I-map for H. Then G is necessarily chordal.

MNs to BNs

- Turning a MN to a BN requires triangulation: adding enough edges to a graph to make it chordal.
- Leads to the loss of independence information
- When converting from BN→MN, the (moralizing) edges added are in some sense implicitly there (ie. each factor in the BN involves a node and its parents)
- When converting from MN→BN, we can introduce a large number of edges (via triangulation) which results in very large cliques

Chordal Graphs

25

Chordal Graphs

Let H be a chordal Markov network. Then there is a Bayesian network G which is a perfect map for H ie. $I(H) = I(G)$

We will introduce concepts to help us sketch out a proof for this.

26

Chordal Graphs

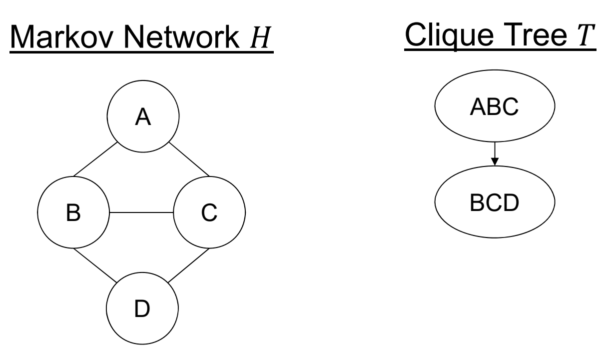
Basic idea:

- Show that any connected chordal Markov network H can be decomposed into a clique tree (to be defined)
- The clique tree encodes independencies in H
- These independencies can be represented in a Bayesian network G

27

Chordal Graphs

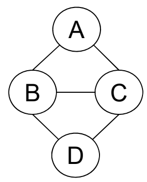
- Let H be a connected undirected graph
- Let $\boldsymbol{C}_1, \dots, \boldsymbol{C}_k$ be the set of maximal cliques in H
- Let T be any tree-structured graph whose nodes correspond to the maximal cliques $\boldsymbol{C}_1, \dots, \boldsymbol{C}_k$



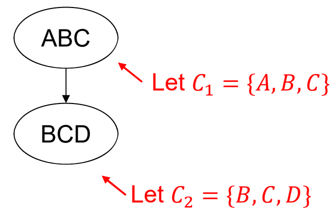
28

Chordal Graphs

Markov Network H



Clique Tree T



- Define the Sepset $S_{1,2} = C_1 \cap C_2$ (Note: the two cliques must be connected by an edge)
- Let $\boldsymbol{W}_{<(1,2)} = \{A\}$ be all of the variables that appear in any clique on the $\boldsymbol{C}_1$ side of the edge
- Let $\boldsymbol{W}_{<(2,1)} = \{D\}$ be all of the variables that appear in any clique on the $\boldsymbol{C}_2$ side of the edge

29

Chordal Graphs

We say that a tree T is a clique tree for H if:

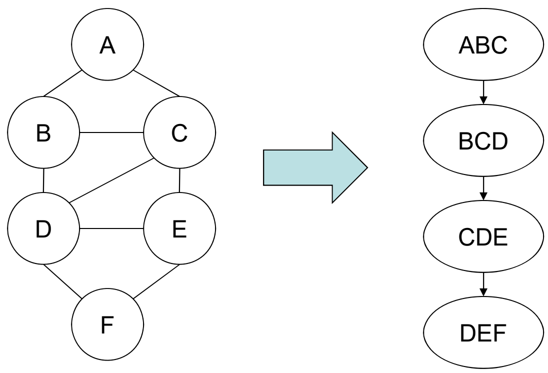
- Each node corresponds to a clique in H, and each maximal clique in H is a node in T
- Each sepset $S_{i,j}$ separates $W_{<(i,j)}$ and $W_{<(j,i)}$ in H
- Note: $W_{<(i,j)} \perp W_{<(j,i)} | \, S_{i,j}$

Theorem 4.12: Every undirected chordal graph H has a clique tree T.
[See book for proof by induction]

30

Chordal Graphs

How do you form a clique tree from bigger Markov networks?



Pick a clique as the root. Then pick an ordering of the cliques that go out from the root (called a topological ordering)

Chordal Graphs

Theorem 4.13: Let H be a chordal Markov network. Then there is a Bayesian network G such that $I(H) = I(G)$.

The process [see full proof in book]:

- Build clique tree T from H
- Call Build-Minimal-I-Map to get Bayesian network G
- G (moralized Baysian network) has the same edges as H due to sepset property $W_{<(i,j)} \perp W_{<(j,i)} | S_{i,j}$