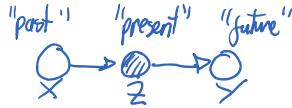


today: DGM (cont.)  
UGM

conditional indep. in DGM

basic 3 nodes graph

1) Markov chain

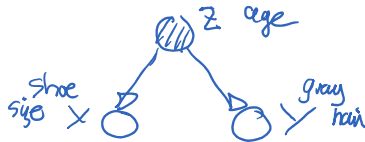


here  $Y \perp\!\!\!\perp X \mid Z$   
but  $X \not\perp\!\!\!\perp Y$   
(for some  $p \in \mathcal{P}(G)$ )

$$p(x, y, z) = p(y|z)p(z|x)p(x)$$

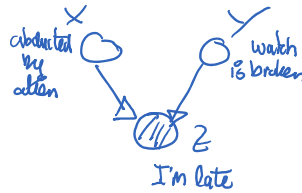
(exercise)  $\Rightarrow p(x, y|z) = p(x|z)p(y|z)$

2) latent cause  
(hidden variable)



same statements as above

3) explaining away / competing effect  
(V structure)



$X \perp\!\!\!\perp Y$   
but  $X \not\perp\!\!\!\perp Y \mid Z$   
(for some  $p$ )

non-monotonic property of conditioning

$$p(\text{alien}) \text{ tiny}$$

$$p(\text{alien} \mid \text{late}) > p(\text{alien})$$

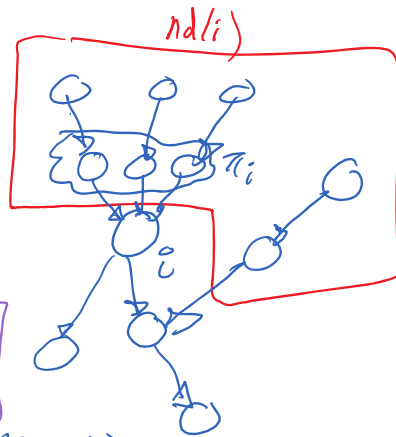
$$p(\text{alien} \mid \text{late, watch broken}) < p(\text{alien} \mid \text{late})$$

more cond. indep. statements in DGM

let  $nd(i) \triangleq \{j \neq i : \text{no path from } i \text{ to } j\}$   
"non-descendants" of  $i$

prop:  $p \in \mathcal{P}(G) \Leftrightarrow X_i \perp\!\!\!\perp X_{nd(i)} \mid X_{\pi_i}, \forall i$

(by decomposition:  $\Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{\pi_i}, \forall j \in nd(i)$ )



proof:

$\Rightarrow$  key point: let  $i$  be fixed, then  $\exists$  a top. ordering of  $G$  s.t.

$nd(i)$  are exactly before  $i$  i.e.  $(nd(i), i, \text{descendant}(i))$

pluck all these?

$$\Rightarrow p(x_i, x_{nd(i)}) = p(x_i | x_{\pi_i}) \prod_{j \in nd(i)} p(x_j | x_{\pi_j})$$

$$p(x_i | \mathcal{I}_{nd(i)}) = \frac{p(x_i, \mathcal{I}_{nd(i)})}{p(\mathcal{I}_{nd(i)})} = \frac{p(x_i | \mathcal{I}_{\pi(i)}) \prod_{j \in nd(i)} p(x_j | \mathcal{I}_{\pi(j)})}{\sum_{x_i} p(x_i | \mathcal{I}_{\pi(i)}) \prod_{j \in nd(i)} p(x_j | \mathcal{I}_{\pi(j)})} = p(x_i | \mathcal{I}_{\pi(i)}) //$$

no  $x_i$  here

$\Leftarrow$ ) (suppose  $p$  satisfies all these cond. indep. statements)

let  $1:n$  be (WLOG) a top. sort. of  $\mathcal{G}$

then  $1:i-1 \subseteq nd(i) \forall i$  [by top. sort property]

$\Rightarrow X_i \perp\!\!\!\perp X_{1:i-1} | \mathcal{I}_{\pi(i)}$  (by decomposition)

$$p(x_V) = \prod_{i=1}^n p(x_i | x_{1:i-1}) \text{ (by chain rule)}$$

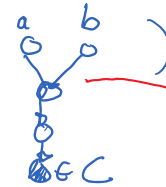
$$= \prod_{i=1}^n p(x_i | \mathcal{I}_{\pi(i)}) \text{ (by cond. indep.)}$$

$$\Rightarrow p \in \mathcal{J}(\mathcal{G}) //$$

other cond. indep. statements?

chain from  $a$  to  $b$ : just any (undirected) path in graph from  $a$  to  $b$

d-separation: (intuition: do not want to have this situation  
 $\rightarrow$  set of observed P.V. conditioned on



def: set  $A$  &  $B$  are said to be d-separated by  $C$  (in  $\mathcal{G}$ )

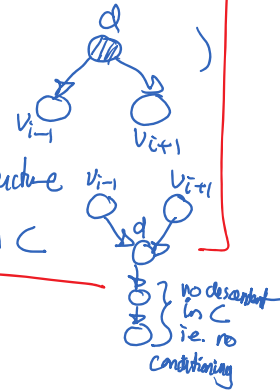
iff all chains from  $a \in A$  to  $b \in B$  are "blocked" given  $C$

where a chain from  $a$  to  $b$  is "blocked" given  $C$  at node  $d$

if a) either  $d \in C$  and  $(v_{i-1}, d, v_{i+1})$  is not a v-structure



b)  $d \notin C$  and  $(v_{i-1}, d, v_{i+1})$  is a v-structure and no descendant of  $d$  is in  $C$



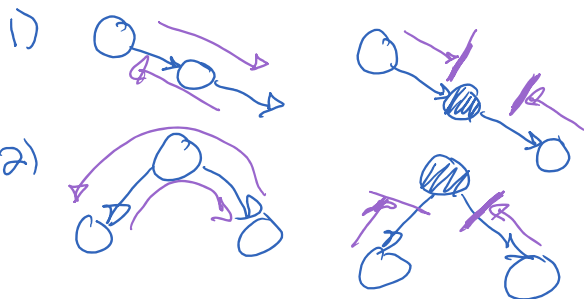
prop:  $p \in \mathcal{J}(\mathcal{G}) \Leftrightarrow X_A \perp\!\!\!\perp X_B | X_C \forall A, B, C \subseteq V$   
 st.  $A$  &  $B$  are d-separated by  $C$

these are all cond. indep. statements which hold  $\forall p \in \mathcal{J}(\mathcal{G})$

"Bayes ball" alg.

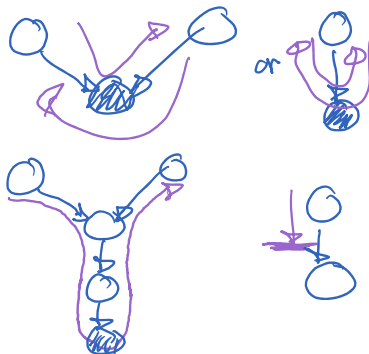
"intuitive" alg. to check d-separation  
 rules balls/chains being blocked

a) in d-separation



b)

3) V-structure



see Alg. 3.1  
in K&F book

16h05

more properties of DGM

inclusion:  $E \subseteq E' \Rightarrow \mathcal{J}(G) \subseteq \mathcal{J}(G')$

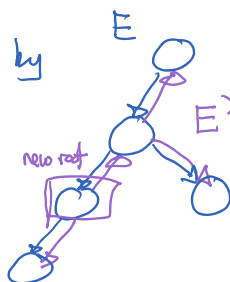
reversal: if  $G$  is a directed tree (or a forest)

$\hookrightarrow \pi_i \perp \pi_j \nVdash i$  i.e. there is no v-structure

Let  $E'$  be another directed tree  
(with same undirected edges) by  
choosing a different root

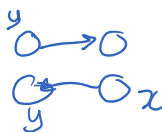
(reverse some edges while keeping  
a directed tree)

$\Rightarrow \mathcal{J}(G) = \mathcal{J}(G')$



rephrasing: all directed trees from  
an undirected tree give  
same DGM

⊛ why direction of  
edges are not causal in DGM



marginalization:

• marginalizing a leaf node  $n$  gives us a smaller DGM

let  $S = \{ q \text{ is a dist. on } \mathcal{X}_{1:n-1} \text{ s.t. } q(\mathcal{X}_{1:n-1}) = \sum_{\mathcal{X}_n} p(\mathcal{X}_{1:n}) \}$   
for some  $p \in \mathcal{J}(G)$

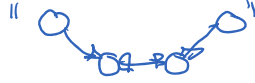
then  $S = \mathcal{J}(G')$  where  $G'$  is  $G$  with leaf  $n$  plucked (removed)

• not true for all marginalizations



↳ get a set of dist.  $\neq$  to any  $S(G)$  for any  $G$

i.e. Marginalization is not a "closed" operation for DGMs



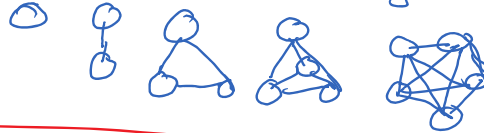
Undirected GM (UGM) (aka. as Markov random field or Markov network)

let  $G=(V,E)$  be an undirected graph

let  $\mathcal{C}$  be the set of cliques of  $G$

clique  $\triangleq$  fully connected set of nodes

(if  $C \in \mathcal{C} \Leftrightarrow \forall_{i,j \in C, i \neq j} (i,j) \in E$ )



UGM associated with  $G$

$f(G) \triangleq \{ p \text{ is a dist. over } X_V : p(x_V) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \}$

for some "potentials"  $\psi_C: \Omega_{X_C} \rightarrow \mathbb{R}_+$

and  $Z \triangleq \sum_{x_V} \left( \prod_{C \in \mathcal{C}} \psi_C(x_C) \right)$

normalization constant

"partition function"

notes:

• unlike a DGM  $\psi_C(x_C)$  is not directly related to  $p(x_C)$

[ in DGM could think of  $C$  as  $(i, \pi_i)$  ]  
 $\psi_C(x_C) = p(x_i | x_{\pi_i})$

• can rescale any potential without changing the joint

i.e.  $\psi_C^{new}(x_C) \triangleq \psi_C^{old}(x_C) \cdot A$  for some  $A > 0$

(unlike in a DGM where  $f_i$ 's were uniquely defined from  $p$ )

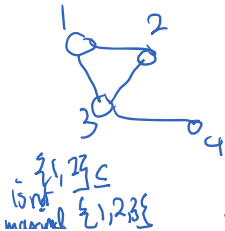
→ potentials are not unique

• it is sufficient to consider  $\mathcal{C}_{max}$ , the set of maximal cliques

e.g.  $C' \subseteq C$

redefine  $\psi_C^{new}(x_C) \triangleq \psi_C^{old}(x_C) \cdot \psi_{C'}^{old}(x_{C'})$

(if  $C'$  belongs to more than one  $C$ , just choose one)



maximal clique

[ we'll see later sometimes convenient to consider "over parametrization" of trees using both  $\psi_i(x_i)$  and  $\psi_{ij}(x_i, x_j)$  ]

$\psi_i$  and  $\psi_{ij}$  which one for  $\psi_j$

properties of UGM:

- as before  $E \subseteq E' \Rightarrow \mathcal{J}(G) \subseteq \mathcal{J}(G')$
- $E = \emptyset \Rightarrow \mathcal{J}(G) = \text{set of fully factorized dist.}$
- $E = \text{all pairs (i.e. } V \text{ is just a big clique in } G) \Rightarrow \mathcal{J}(G) = \text{all dist. on } x_V$

• if  $\psi_c(x_c) > 0 \forall x_c, \forall c$

then can write  $p(x_V) = \exp\left(\sum_{c \in \mathcal{C}} \log \psi_c(x_c) - \log Z\right)$   $\dim = |X_c|$

$\langle \Theta_c, T_c(x_c) \rangle$  where  $T_c(x_c) = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$  " $x_c$ "

$\Theta_c(x_c) = \log \psi_c(x_c)$

$\sum_c \sum_{x_c} \Theta_c(x_c) T_c(x_c)$

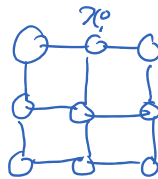
physics link  $\rightarrow$  negative energy fact.

eg. Ising model in physics

$x_i \in \{-1, 1\}$

node potential  $E_i = \log \psi_i(x_i = 1)$

edge potential  $E_{ij} = \log \psi_{ij}(x_i = 1, x_j = 1)$



other example: social network