Probabilistic Graphical Models.

Graph Separation $\Longleftrightarrow$ $I(G) = \bigcap_{P \text{ factorizes wrt } G} I(P)$

$\Uparrow$

Graphs G $\Rightarrow$ Factorization $P(x) = \Pi$ $\Rightarrow$ Conditional Independence $I(P)$

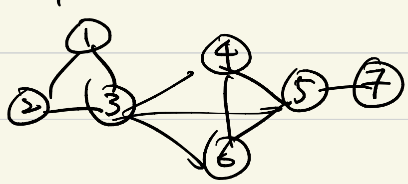
* Undirected Graphical Models

Def. Undirected Graph $G = (V, E)$

$\quad \llcorner (i,j) = (j,i)$

- a set of nodes $C = \{i_1, i_2, \dots, i_k\}$ is a clique in G if all pairs are connected in G.

- a clique C is maximal clique if you cannot add any more nodes without breaking clique.



$\{1, 2, 3\}$
$\{3, 4, 5, 6\}$
$\{5, 7\}$

- $\mathcal{C}$ to denote set of all maximal cliques.

Def. Undirected Graphical Model on $G = (V, E)$ is a family of distributions on $X = [X_1 \dots X_n]$ that factorizes as

// Markov Random Fields (MRF)

$$P(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} f_c(x_c)$$

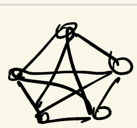
$f_c : \mathcal{X}^{|c|} \to \mathbb{R}_+$ factors or Compatibility functions

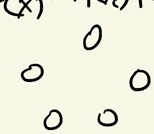
$Z \triangleq \sum_{x \in \mathcal{X}^n} \left\{ \prod_{c \in \mathcal{C}} f_c(x_c) \right\}$ called a Partition Function.

unstructured $\longleftrightarrow$ Structured

$P(x)$ distribution: $|I(P)|$ is small $\longleftrightarrow$ $|I(P)|$ large $\quad P(x) = P(x_1)P(x_2)\dots$

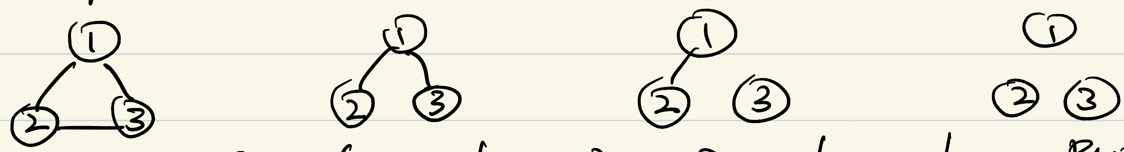
graph: dense $\longleftrightarrow$ sparse





Warm up examples >

G:



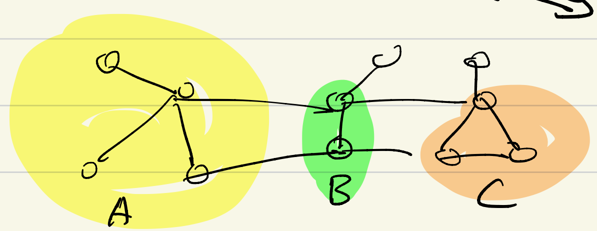
$P(x)$ | $P(x) \propto f_{12}(x_1,x_2)\, f_{13}(x_1,x_3)$ | $P(x) \propto f_{12}(x_1,x_2)\, f_3(x_3)$ | $P(x) = \frac{1}{Z} f_1(x_1)\, f_2(x_2)\, f_3(x_3)$

$\downarrow$ HW1

$X_2 \perp\!\!\!\perp X_3 \mid X_1$ | $(X_1,X_2) \perp\!\!\!\perp X_3$ | X_1, X_2, X_3 mutually independent.

Def. [Graph Separation]
$B \subset V$ separates $A, C \subset V$ if every path from any $a \in A$ to $c \in C$ passes through B.

$\Longleftrightarrow$ *I will casually write
If $A - B - C$



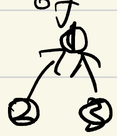
Def. [Global Markov Property (G)]
A distribution $P(x)$ satisfies Global Markov Property (G) if for any $A-B-C$ we have

$$X_A \perp\!\!\!\perp X_C \mid X_B$$

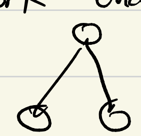
punch line: factorization $\Rightarrow$ (G)

$\exists p$
$I(G) \subset I(p)$
strict for all G

Q. give an example of Bayesin Network that cannot be represented by a MRF.



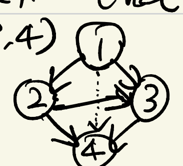
BN
$X_2 \perp\!\!\!\perp X_3$

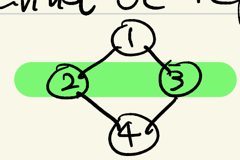


MRF
$X_2 \perp\!\!\!\perp X_3 \mid X_1$

Q. give an example of MRF that cannot be represented by a BN.

BN
$(1,2,3,4)$





MRF $I(G)$:
$X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_3)$
$X_2 \perp\!\!\!\perp X_3 \mid (X_1, X_4)$

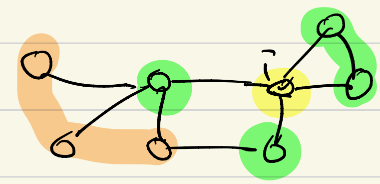
Def. [local Markov Property (L)]

$P(x)$ satisfies (L) w.r.t G if

$$X_i \perp\!\!\!\perp X_{rest} \mid X_{\partial i}$$

∂i = neighborhood of node $i \triangleq \{j \in V \mid (i,j) \in E\}$

$rest = V \setminus (\{i\} \cup \partial i)$

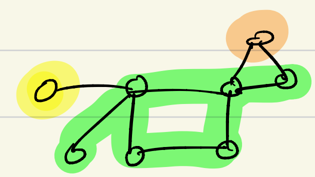


Def [Pairwise Markov Property (P)]

$P(x)$ satisfies (P) w.r.t G if

$X_i \perp\!\!\!\perp X_j \mid X_{rest}$ for all i,j not connected

$rest = V \setminus \{i,j\}$



claim: (G) $\Rightarrow$ (L) $\Rightarrow$ (P)

proof 1: (G) $\Rightarrow$ (L)

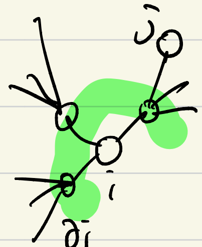
follows from $i - \partial i - rest$ are separated in Graph.

(G): $X_A \perp\!\!\!\perp X_C \mid X_B$

(L): $X_i \perp\!\!\!\perp X_{rest} \mid X_{\partial i}$

(P): $X_i \perp\!\!\!\perp X_j \mid X_{rest}$

proof 2: (L) $\Rightarrow$ (P)



$$X_i \perp\!\!\!\perp X_{V \setminus (\{i\} \cup \partial i)} \mid X_{\partial i}$$

any j not in ∂i

$$\overset{(a)}{\Longrightarrow} X_i \perp\!\!\!\perp X_{V \setminus (\{i\} \cup \partial i)} \mid X_{V \setminus \{i,j\}}$$

used fact that $j \notin \partial i$

$$\overset{(b)}{\Longrightarrow} X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}} \Longleftrightarrow (P)$$

(a) follows from $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid (Z, h(Y))$ for any $h(\cdot)$

(b) follows from $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp h(Y) \mid Z$ for any $h(\cdot)$

$\uparrow$ Data Processing Inequality. $X - \boxed{Z} - Y - h(Y)$

$(G) \Rightarrow (L) \Rightarrow (P)$ PMF: $P(x) > 0 \quad \forall x = [x_1 \cdots x_n] \in \mathcal{X}^n$

Claim: If $P(x) > 0 \ \forall x$, then $(P) \Rightarrow (G)$

> **Lemma [intersection lemma]**
> For $P(x) > 0$, if $\left.\begin{array}{l} X_A \perp\!\!\!\perp X_B \mid (X_C, X_D) \\ X_A \perp\!\!\!\perp X_C \mid (X_B X_D) \end{array}\right] \Rightarrow X_A \perp\!\!\!\perp (X_B X_C) \mid X_D$

proof of claim: $X_i \perp\!\!\!\perp X_j \mid X_{rest} \Rightarrow X_A \perp\!\!\!\perp X_C \mid X_B$

show by induction, on all B of size $|B| = n-2, n-3, \ldots$

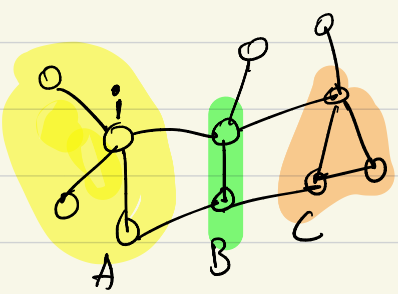
① initially, $|B| = n-2$. $A = \{i\}, C = \{j\}$
true by (P)

② suppose by induction that (G) holds for all $|B| \geq s$.
we will show for B of size $|B| = s-1 < n-2$
suppose w.l.o.g. $|A| \geq 2$.

$A - B - C$ separated in Graph.

$\Rightarrow \underbrace{A \setminus \{i\}}_{\tilde{A}} - \underbrace{B \cup \{i\}}_{\substack{\text{size } s+1 \\ \tilde{B}}} - C$ also separated

$X_C \perp\!\!\!\perp X_{A \setminus \{i\}} \mid X_{B \cup \{i\}}$ (1)



$\Rightarrow \underbrace{\{i\}}_{\tilde{A}} — \underbrace{B \cup A \setminus \{i\}}_{\substack{\tilde{B} \\ \text{size} > s}} - C$ also separated (2)

$X_C \perp\!\!\!\perp X_i \mid (X_B, X_{A \setminus \{i\}})$

Apply intersection lemma: (1) & (2)

$X_C \perp\!\!\!\perp X_A \mid X_B.$

$\forall A, B, C$ s.t. $A - B - C$ separated & $|B| = s$. □

* If $P(x) \not> 0$, then we can construct a counterexample s.t. $(G) \not\Leftarrow (P)$. [HW1]

Def. [Factorization (F)]

we say $P(x)$ factorizes w.r.t G if

$$P(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} f_c(x_c)$$

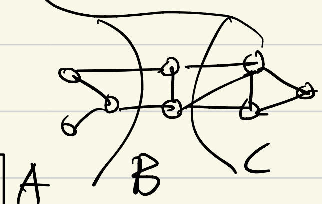
" maximal clique set

Claim: (F) $\Rightarrow$ (G)

proof: $P(x) \propto \prod_{c \in \mathcal{C}} f_c(x_c)$. & A — B — C

$\propto F_{AB}(x_A, x_B) \cdot F_{BC}(x_B, x_C)$ $\Leftarrow$ no factor has both $a \in A, c \in C$

H.W 1 $\longrightarrow$ $x_A \perp\!\!\!\perp x_C \mid x_B$



$$\boxed{\begin{array}{l}\text{Claim: (G)} \Rightarrow \text{(F)}, \text{ if } P(x) > 0 \\ \text{[Hammerseley-Clifford theorem]}\end{array}}$$

implication: for $P(x) > 0$, (F) $\Longleftrightarrow$ (G) $\Longleftrightarrow$ (L) $\Longleftrightarrow$ (P)

Counterexample: If $P(x) \not> 0$, then (G) $\not\Rightarrow$ (F)