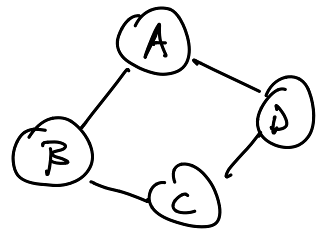
Lecture 4: UNDIRECTED GRAPHICAL MODELS

[0] Motivation

→ Some problems may require different independencies than those encoded by DAGs / d-separation

Ex: (K&F, Chap. 4) "Misconception" example

4 students A, B, C, D, work in pairs each may have a misconception about a topic left vague by professor.

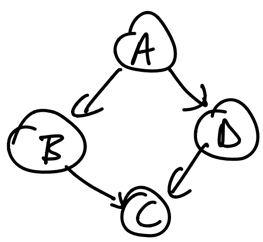


- Desired independencies: $B \perp D \mid \{A, C\}$

  $A \perp C \mid \{B, D\}$

- Can a directed model represent these? <u>NO</u>

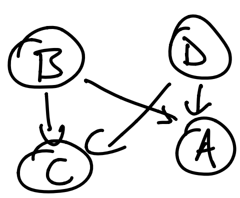
e.g.



$\Rightarrow$ $A \perp C \mid \{B, D\}$ ✓

$B \perp D \mid A$!!

$B \not\perp D \mid \{A, C\}$!

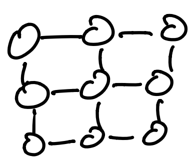


$\Rightarrow$ $A \perp C \mid \{B, D\}$ ✓

$B \perp D$!!

$\Rightarrow$ Need a different language to encode such C.I.s

Ex: pairwise interactions in Ising models in physics (magnetic forces)



[1.] Markov Networks and Gibbs Distributions

<u>Idea:</u> Represent $p(x)$ as product of Factors $\phi(x_C)$
$C \subseteq V$ is a set of variables

<u>Def:</u> (Gibbs distribution)
$p_\phi(x)$ is a Gibbs distribution if it take the form

$$p_\phi(x) = \frac{1}{Z} \phi_1(x_{C_1}) \cdots \phi_m(x_{C_m})$$

for some factors $\phi = \{\phi_i(x_{C_i})\}_i$ on variables $C_i \subseteq V$

$Z := \sum_{x_1,\dots,x_d} \phi_1(x_{C_1}) \cdots \phi_m(x_{C_m})$ is the partition Function (normalizing constant)

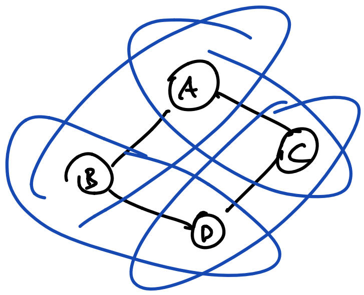
<u>Note:</u>
- A distribution $p(x)$ that factorizes on a DAG can also be seen as a product of factors $\phi_i(x_i, x_{pa(i)}) = p(x_i \mid x_{pa(i)})$, but with the <u>key distinction</u> that each factor ϕ_i needs to be <u>normalized</u> s.t. $\sum_{x_i} \phi_i(x_i, x_{pa(i)}) = 1$ for any $x_{pa(i)}$
- For $x \in \{0,1\}^m$, the # of parameters in such a Gibbs distr may grow exponentially in $\max_i |C_i|$, which can be $\ll |V| = d$

Def: (Markov Network factorization)

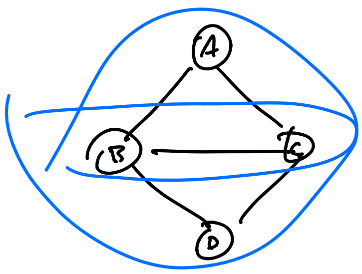
Let $H=(V,E)$ be an undirected graph.

A distr. $P_\Phi(x)$ with factors $\Phi=\{\phi_i(x_{C_i})\}_i$ factorizes over the Markov Network H if each C_i is a complete subgraph (clique) of H

Ex:



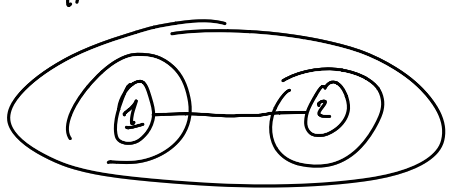
$$P_\Phi(x)=\frac{1}{Z}\phi_1(x_A,x_B)\,\phi_2(x_B,x_D)\,\phi_3(x_C,x_D)\,\phi_4(x_A,x_C)$$

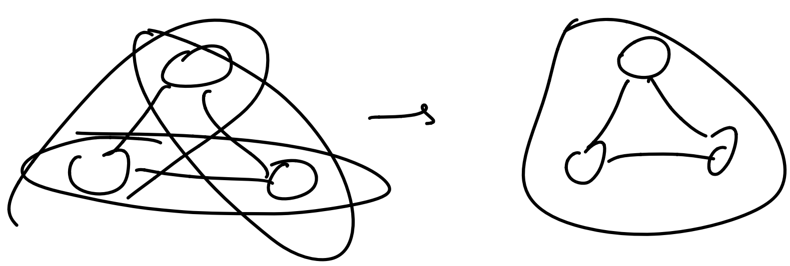


$$P_\Phi(x)=\frac{1}{Z}\phi_1(x_A,x_B,x_C)\,\phi_2(x_B,x_C,x_D)$$

Note: • We can always reduce # factors to only include factors on maximal cliques

e.g. $\phi_1(x_1)\,\phi_2(x_2)\,\phi_{12}(x_1,x_2) \dashrightarrow \Phi(x_1,x_2)$





- However, maximal cliques may obscure some underlying structure, and increase complexity in the parameterization (e.g. hide pairwise potentials, use less structured parameterization)

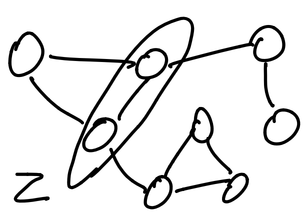
2. Markov Network independencies

- Global independencies.

Q: What Cond. Indep.s does a Markov Net encode?
Links with factorization as in directed case?

Def: Let $Z \subseteq V$ and H a Markov net. structure over V
A path $i_1 - i_2 - \cdots - i_k$ is active given Z
if none of the i_j, $j=1..k$, are in Z.

Def (separation)
A set of nodes Z separates X and Y in H if there is no active path from any $i \in X$ to $j \in Y$ given Z.
Denoted $sep_H(X; Y|Z)$
We write $\mathcal{I}(H) = \{(X \perp Y | Z);\ sep_H(X;Y/Z)\}$



Q : Do we have H I-map for $P \iff P$ factorizes according to H?
(i.e. $I(H) \subseteq I(P)$)

Theorem (factorization $\Rightarrow$ I-map)
Let $p(x)$ be some distribution, and H be a Markov Net.
If $p(x)$ is a Gibbs distribution that factorizes over H,
then H is an I-map for P.

proof :

Let A, B, C s.t. $sep_H(A; B | C)$

Assume $A \cup B \cup C = V$, we want to show

$x_A \perp x_B | x_C$ in P

(if not, consider $\hat{A} = A \cup U_A$, $\hat{B} = B \cup U_B$
with $\hat{A} \cup \hat{B} \cup C = V$ and $sep_H(\hat{A}; \hat{B} | C)$
deduce $x_A \perp x_B | x_C$ from $x_{\hat{A}} \perp x_{\hat{B}} | x_C$)

We can write :

$$p(x) = \frac{1}{Z} \phi_1(x_A, x_C) \phi_2(x_B, x_C)$$

$$= \tilde{\phi}_1(x_A | x_C) \tilde{\phi}_2(x_B | x_C) \cdot \psi(x_C)$$

with $\tilde{\phi}_1, \tilde{\phi}_2$ normalized

It follows that ψ is also normalized

$\Rightarrow x_A \perp x_B | x_C$

$\blacksquare$

Q: Do we have the other direction?

No! in general, Yes! for positive distributions $p > 0$

Theorem: (I-map $\Rightarrow$ factorization)
Hammersley-Clifford

Let $p > 0$. If H is an I-map for P, then P is a Gibbs distribution that factorizes over H

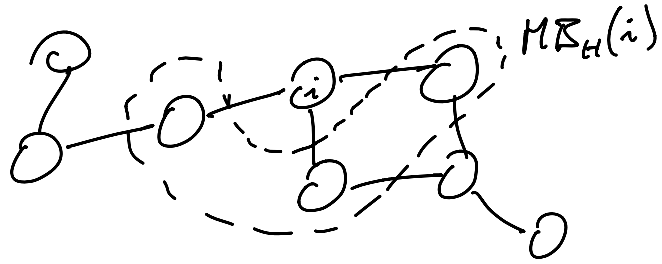
■ Local independencies

Def: (Markov blanket)

$$MB_H(i) = \{j : \{i,j\} \in E\}$$ (neighbors of i)

(Local independencies)

$$\mathcal{I}_\ell(H) = \{(x_i \perp x_{V\setminus(\{i\}\cup MB_H(i))} \mid x_{MB_H(i)})\}$$



Prop: If p satisfies $I(H)$, then it satisfies $I_\ell(H)$

indeed $I_\ell(H) \subseteq I(H)$

Prop: If $\underline{\underline{p>0}}$, and p satisfies $I_\ell(H)$, then it satisfies $I(H)$

In fact, an even smaller set of C.Is suffices if $p>0$:

pairwise indeps:

$$I_p = \{(x_i \perp x_j \mid x_{V\setminus\{i,j\}}),\ \{i,j\} \notin E\}$$

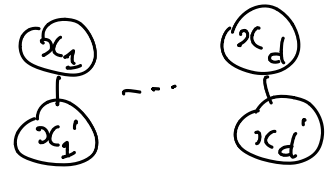
$\Rightarrow$ For positive distributions,

local indeps (Markov blanket) $\iff$ global indeps

Remark: For non-positive p, this is no-longer true

e.g. define $x_i' = x_i$ w.p. 1. (non-positive p!)

We have $x_i' \perp x_j \mid x_i \quad \forall i \neq j$

Hence $I_\ell(H) \subseteq I(p)$ where H: 

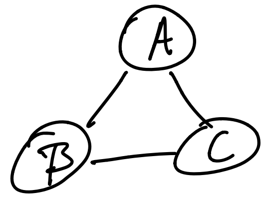
Even though p may have many more (global) indep.

3. Factor graphs, moralization

- Factor graphs

→ Undirected graphs may lose some information about factors

Ex:



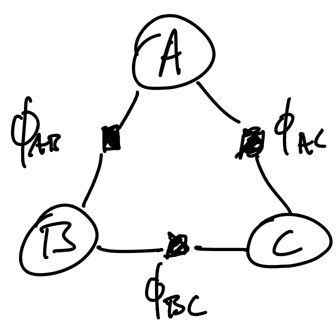
do we have:

$$p(x) \propto \phi(x_A, x_B) \cdot \phi(x_B, x_C) \cdot \phi(x_A, x_C)$$

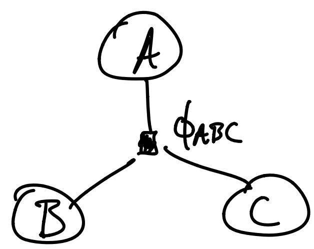
or

$$p(x) \propto \phi(x_A, x_B, x_C) \quad ?$$

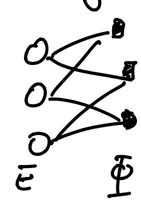
→ A factor graph may help recover this information by encoding factors as nodes



vs.



Bipartite graph: edges are only between nodes - factors (no node-node or factor-factor)



$\mathcal{F} = (V, \Phi, E)$ with $E \subseteq V \times \Phi$

Def: p factorizes over $\mathcal{F}$ if it can be written

$$p(x) = \frac{1}{Z} \prod_{\phi \in \Phi} \phi(x_{\mathcal{N}(\phi)})$$

, with $\mathcal{N}(\phi) = \{i : (i, \phi) \in E\}$

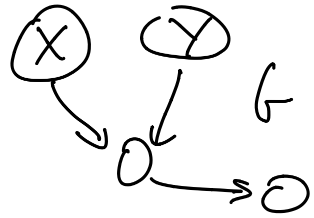
From directed to undirected

Def. (moralization)

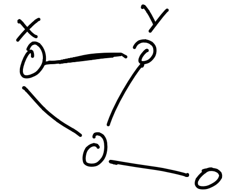
The Moral graph $M(G)$ of a Bayes net G is an undirected graph with edges $X - Y$ if:

- $X \to Y$
- $Y \to X$
- $X \rightarrow Z \leftarrow Y$ (parents in a v-structure)

Ex:



$G \implies M(G)$



Note: the independence $X \perp Y$ is lost!

Prop: We have $I(M(G)) \subseteq I(G)$

If a directed graph G is "moral" (i.e.: no "immoralities")

then $I(M(G)) = I(G)$

Remarks:

- Moralization can be used to show "soundness" of d-sep (i.e. that p factorizes on $G \Rightarrow G$ is an I-map for P)

- Going from undirected to directed is also possible, but more complicated (involves "chordal" graphs)

- In directed models, v-structures lead to more intricate dependencies (e.g. explaining away) when observing new variables.
  Not the case with undirected (observing $\Rightarrow$ remove dependencies only)