ECE 175B: Probabilistic Reasoning and Graphical Models
Lecture 11: Expressive Power of Graphical Models

Florian Meyer & Ken Kreutz-Delgado

Electrical and Computer Engineering Department
University of California San Diego

UC San Diego
JACOBS SCHOOL OF ENGINEERING

Roughly Two Classes of Graphical Models.

- Semantically Interpretable.
  — Bayesian / Belief Networks (BN)
  informs directed behavior / informative-flow.
  — Markov Networks (MN / MRF)
  informs cooperative behavior
  — Chained / Mixed.
  informs mixed directed / cooperative behavior
  — Influential diagrams / Markov Decision Nets
  informs consequences when decisions / actions have costs.
- Computationally Tractable
  — Undirected graphs (e.g. serial chains)
  — Factor Graphs
  — Junction trees

Basic Conditional Independences (CIs) Statements of BNs

$$P(\mathcal{X}) = \prod_{j=1}^{N} p(x_j \mid \underbrace{x_{j-1}, \dots, x_1}_{\text{Prior to}(x_j) = \text{Predecessors}(x_j) = \text{Pred}(x_j)}) = \prod_{j=1}^{N} p(x_j \mid Pa(x_j)) \quad (*)$$

where $Pa(x_j)$ is the smallest set such that

$$\boxed{x_j \perp\!\!\!\perp \text{Pred}(x_j) \mid Pa(x_j)} \quad \text{for } Pa(x_j) \subset \text{pred}(x_j) \quad (**)$$

- Note that $\{x_1, \dots, x_N\}$ is one of the $N!$ possible orderings of $\mathcal{X}$. Given an ordering, the factorization $(*)$ is <u>unique</u> when P is <u>positive</u> $\Leftrightarrow P(\mathcal{X}) > 0, \forall \mathcal{X}$.

- Note that $(**)$ makes sense because $X \perp\!\!\!\perp Z \mid Z$, which is true because $X \perp\!\!\!\perp Y \mid Z$ i.f.f. $P(X \mid Y, Z) = P(X \mid Z)$, which is true for $Y = Z$.

$(**)$ is the $\boxed{\text{Basic Markov Property}}$ for a BN.

Basic CI Statements for BNs - Cont'd.

- By analyzing the properties of collider and noncollider nodes on a path, we earlier show the $\boxed{\text{Global Markov Property}}$ of a BN $\mathcal{G} = (V, E)$.

For <u>disjoint</u> $X, Y, Z \subset \mathcal{X}, \longleftrightarrow V$

$$\boxed{\underbrace{\langle X \mid Z \mid Y \rangle_d}_{X \text{ and } Y \text{ are d-separated on the BN}} \implies X \perp\!\!\!\perp Y \mid Z} \qquad (*)$$

- Define the non-descendents of x_j by $\text{nondesc}(x_j) = \mathcal{X} \setminus (\text{desc}(x_j) \cup \{x_j\})$ we have the $\boxed{\text{Parental Markov Property}}$ of a BN:

$$\boxed{x_j \perp\!\!\!\perp \text{nondesc}(x_j) \mid Pa(x_j)} \qquad (**)$$

Basic CI statements for BNs - Cont'd.

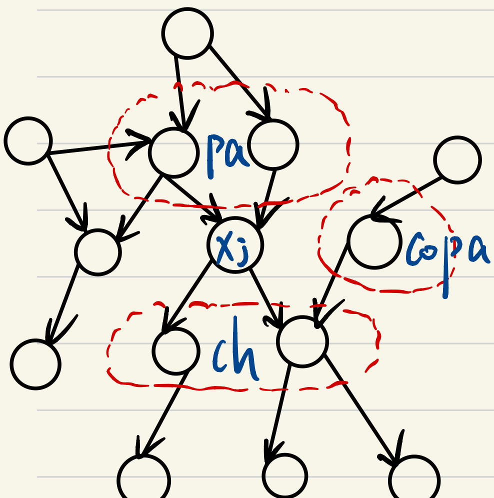
- Define some sets of $x_j \in \mathcal{X}$ in a BN:

children: $ch(x_j) = \{x \mid x_j \in pa(x)\} \subset \mathcal{X}$

coparents: $copa(x_j) = \{x \in pa(z) \mid z \in ch(x_j),\ x \neq x_j\}$

Markov Blanket: $mb(x_j) = ch(x_j) \cup pa(x_j) \cup copa(x_j)$

closure: $CL(x_j) = mb(x_j) \cup \{x_j\}$ and $CL^c(x_j) = \mathcal{X} \setminus CL(x_j)$



The $\boxed{\text{Local Markov Property}}$ of a BN:

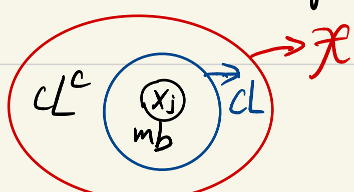
$$\boxed{x_j \perp\!\!\!\perp CL^c(x_j) \mid mb(x_j)}$$

$mb = pa \cup copa \cup ch$

shields x_j from the rest of the BN.

$\downarrow$

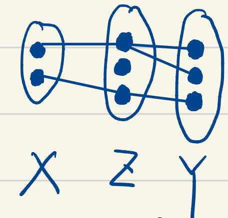
$CL^c(x_j)$



Basic CI statements for MNs.

- $\boxed{\text{Global Markov (GM) Property}}$ For disjoint $X, Y, Z \subset \mathcal{X}$

$$\boxed{\begin{array}{l} \langle X | Z | Y \rangle_d \Longrightarrow X \perp\!\!\!\perp Y | Z \\ \hookrightarrow \text{All paths from } X \text{ to } Y \text{ go through } Z. \end{array}}$$



(*)

- A = adjacency matrix. $A(i,j) = \begin{cases} 1 & \text{; if an edge } (X_i, X_j) \text{ exists} \\ 0 & \text{; otherwise} \end{cases}$
- Define some sets of $X_j \in \mathcal{X}$ in a MN:
  - neighborhood: $nbr(X_j) = \{X_i \mid A(i,j) = 1\}$
  - Markov Blanket: $mb(X_j) \triangleq nbr(X_j)$
  - closure: $CL(X_j) = mb(X_j) \cup \{X_j\}$; $CL^c(X_j) = \mathcal{X} \setminus CL(X_j)$
- $\boxed{\text{Local Markov (LM) Property}}$

$$\boxed{\langle X_j | \underbrace{mb(X_j)}_{nbr(X_j)} | CL^c(X_j) \rangle \Longrightarrow X_j \perp\!\!\!\perp CL^c(X_j) \,|\, \underbrace{mb(X_j)}_{nbr(X_j)}}$$

(**)

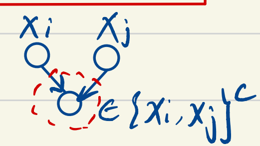
Basic CI Statements for MNs — Cont'd.

- $\boxed{\text{Pairwise Markov (PM) Property}}$ of MNs

  For any two random variables, note $\{X_i, X_j\}^c = \mathcal{X} \setminus \{X_i, X_j\}$,

  $$\boxed{\underbrace{\langle X_i \mid \{X_i, X_j\}^c \mid X_j \rangle}_{\text{true i.f.f. } A(i,j)=0} \Longrightarrow X_i \perp\!\!\!\perp X_j \mid \{X_i, X_j\}^c} \quad (*)$$

  Note that PM property is not true for BNs.



$A(i,j) = 0$ but $X_i \not\!\perp\!\!\!\perp X_j \mid \{X_i, X_j\}^c$

- $\boxed{\text{Factorization (F) Property}}$ of MNs.

  $$\boxed{P(\mathcal{X}) \propto \prod_{\underbrace{C \in \mathcal{C}}_{\text{cliques of MN}}} \phi_C(\mathcal{X}_C) \geq 0 \; ; \quad \phi_C(\mathcal{X}_C) \geq 0} \quad (**)$$

- It's the case that (see Koller) $F \Rightarrow GM \Rightarrow LM \Rightarrow PM$.
- If $\underline{\underline{P > 0}}$, then $PM = PM^+ \Longrightarrow F = F^+$, and all are equivalent.
  (Hammersley–Clifford Theorem)

Basic CI Statements for MNs - Cont'd

- Thus, for $MN^+ = MRF$, when $P(\mathcal{X})$ is a Gibbs distribution,
$$0 < P(\mathcal{X}) = \frac{1}{Z} e^{-\sum_{c \in \mathcal{C}} E_c(\mathcal{X}_c)} \quad ; \quad \phi_c(\mathcal{X}_c) = e^{-E_c(\mathcal{X}_c)} > 0$$

We have P is a Gibbs distribution i.f.f. P has a compatible Globally Markov undirected graph representation with

$$\boxed{F^+ \Longleftrightarrow GM^+ \Longleftrightarrow LM^+ \Longleftrightarrow PM^+} \qquad (*)$$

- Thus, if you can show that a collection $\mathcal{X}$ of random variables are pairwise Markov and positive $(P(\mathcal{X}) > 0)$, then $\mathcal{X}$ satisfies the LM and GM properties and can be modelled by a Gibbs distribution.

The Semantics of Probabilistic Graphical Models (PGMs)

- The semantics of PGMs are given by the CI statements encoded in the graph according to the appropriate graph separation criteria. These are CI statements for $\mathcal{X}$ compatible with the graph.
- A set of CI statements is denoted by $\mathbb{I}$.
- If $\mathbb{I}' \subset \mathbb{I}$, we say that $\mathbb{I}'$ is "faithful" to $\mathbb{I}$.
- If $\mathbb{I}' = \mathbb{I}$, we say that $\mathbb{I}'$ is "perfect" for $\mathbb{I}$.
- Given a distribution $P(\mathcal{X})$, the CI statements true for P are denoted by $\mathbb{I} = \mathbb{I}(P)$.
- Give a graph $\mathcal{G}$ (of any kind), its CI statements are denoted by $\mathbb{I} = \mathbb{I}(\mathcal{G})$.
- If $\mathbb{I}(\mathcal{G}) \subset \mathbb{I}(P)$, we say $\mathcal{G}$ is "faithful" to P.
  If $\mathbb{I}(\mathcal{G}) = \mathbb{I}(P)$, we say $\mathcal{G}$ is "perfect" for P. (can be almost impossible to find)
- If $\mathbb{I}(\mathcal{G}') \subset \mathbb{I}(\mathcal{G})$, we say $\mathcal{G}'$ is "faithful" to $\mathcal{G}$.
  If $\mathbb{I}(\mathcal{G}') = \mathbb{I}(\mathcal{G})$, we say $\mathcal{G}'$ is "perfect" for $\mathcal{G}$.

Perfection and Equivalence of BNs

- Two BNs, G and G', are Markov Equivalent (ME) by definition iff. they are "perfect" for each other, i.e., $\mathbb{I}(G) = \mathbb{I}(G')$
- Therefore, we also call this $\mathbb{I}$-equivalence (IE) ("independence equivalence")
- The connectivity structure of two ME graphs cannot be determined from observational (non-experimental) data alone. Thus we also call two ME graphs observationally Equivalent. (OE)
- Thus, OE = IE = ME = "naturally perfect" for graphs.
- In general, a graph will not be "perfect" for $P(\mathcal{X})$.

  This is because no single type of graph (MN, BN, chain, etc.) can encode all possible CI statements in $\mathbb{I}(P)$, as we have seen.

  If a graph is compatible with P, it will be "faithful", i.e., $\mathbb{I}(G) \subset \mathbb{I}(P)$

Moralization of BNs

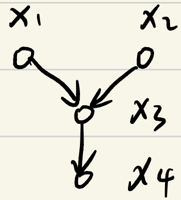
- Consider a distribution $p(\mathcal{X})$ with CI statements $\mathcal{I}(p)$.
- Find a MN for P, call it $\mathcal{N}$. They must be compatible $j \leftrightarrow x_j$ with
  $$p(\mathcal{X}) = \prod_{A \in \mathcal{C}} \phi_A(\mathcal{X}_A), \quad A \in \mathcal{C} = \text{cliques of } \mathcal{N}.$$
- Find a BN for P, call it $\mathcal{G}$. They must be compatible $j \leftrightarrow x_j$ with
  $$p(\mathcal{X}) = \prod_{j=1}^{N} p(x_j \mid pa(x_j))$$
- The moral graph of $\mathcal{G}$ is the U-graph obtained by setting
  $$\phi_{A_j}(\mathcal{X}_{A_j}) = p(x_j \mid pa(x_j)), \quad A_j = \{x_j\} \cup pa(x_j)$$

When we draw this graph, call it $M(\mathcal{G})$. A_j must be a clique, i.e., there are no arrows, and there exist edges between every element, i.e., we "marry" all members of $pa(x_j)$.

Moralization of BNs – Cont'd.

- Example:

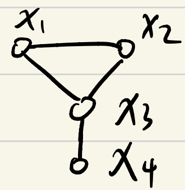
G:



$$P(x_1, x_2, x_3, x_4) = P(x_4 \mid x_3)\, P(x_3 \mid x_1, x_2)\, P(x_2)\, P(x_1)$$

$$\mathbb{I}(G) = \{ \underbrace{x_1 \perp\!\!\!\perp x_2}_{①},\ \underbrace{\{x_1, x_2\} \perp\!\!\!\perp x_4 \mid x_3}_{②} \}$$

$M(G)$:



$$P(x_1, x_2, x_3, x_4) = \phi_{A_1}(x_{A_1})\, \phi_{A_2}(x_{A_2})$$

$A_1 = \{1, 2, 3\}$, $A_2 = \{3, 4\}$ are cliques.

$$\langle \{x_1, x_2\} \mid x_3 \mid x_4 \rangle \Longrightarrow \mathbb{I}(M(G)) = \{ \{x_1, x_2\} \perp\!\!\!\perp x_4 \mid x_3 \}$$

In this case: $\mathbb{I}(M(G)) \subset \mathbb{I}(G)$ but $\mathbb{I}(M(G)) \neq \mathbb{I}(G)$

$\Downarrow$ faithful

$\Downarrow$ Not perfect.