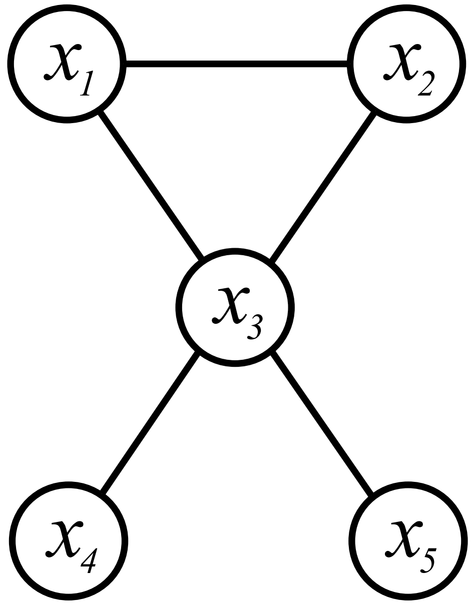
Probabilistic Graphical Models

Brown University CSCI 2950-P, Spring 2013
Prof. Erik Sudderth

Lecture 3:
Undirected Graphical Models

Some figures courtesy Michael Jordan's draft textbook,
An Introduction to Probabilistic Graphical Models

Undirected Graphs



$\mathcal{V} \longrightarrow$ set of N nodes or vertices, $\{1, 2, \ldots, N\}$

$\mathcal{E} \longrightarrow$ set of undirected edges (s,t), or equivalently (t,s), linking pairs of nodes. The *neighbors* of a node are

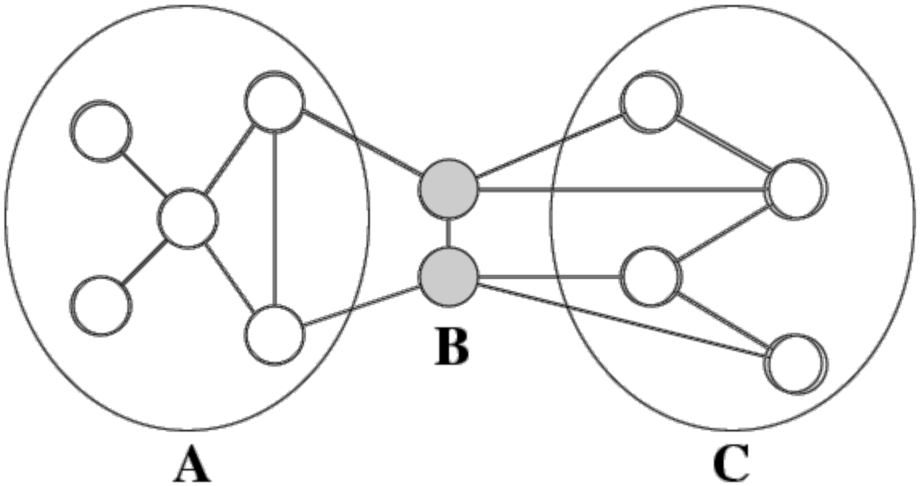
$$\Gamma(t) = \{s \in \mathcal{V} \mid (s,t) \in \mathcal{E}\}$$

$X_s = x_s \longrightarrow$ random variable associated with node s

Undirected Graphical Models

$\mathcal{V}$ ⟶ set of N nodes $\{1, 2, \ldots, N\}$

$\mathcal{E}$ ⟶ set of edges (s, t) connecting nodes $s, t \in \mathcal{V}$



Graph Separation

⟺

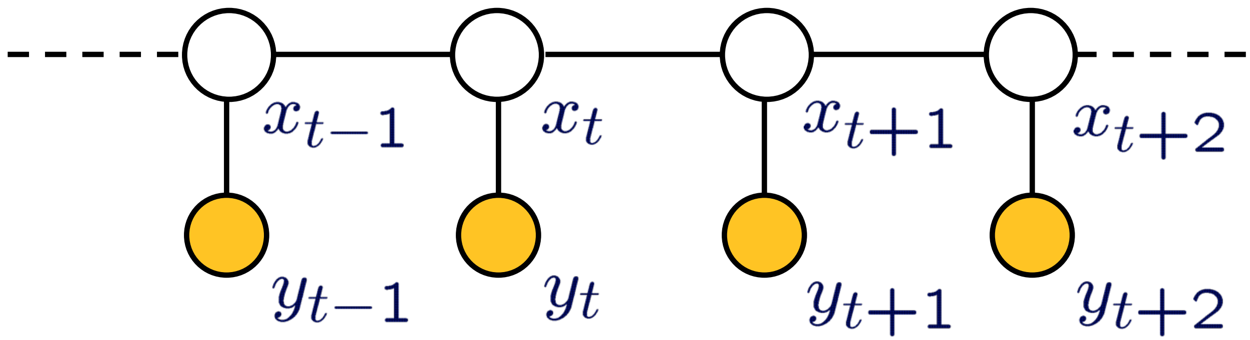
Conditional Independence

$$p(x_A, x_C | x_B) = p(x_A | x_B) p(x_C | x_B)$$

- Simple graph separation, no complexities of directed models.
- This global Markov property implies a local Markov property:

$$p\big(x_i \mid x_{\mathcal{V}\setminus i}\big) = p\big(x_i \mid x_{\Gamma(i)}\big)$$

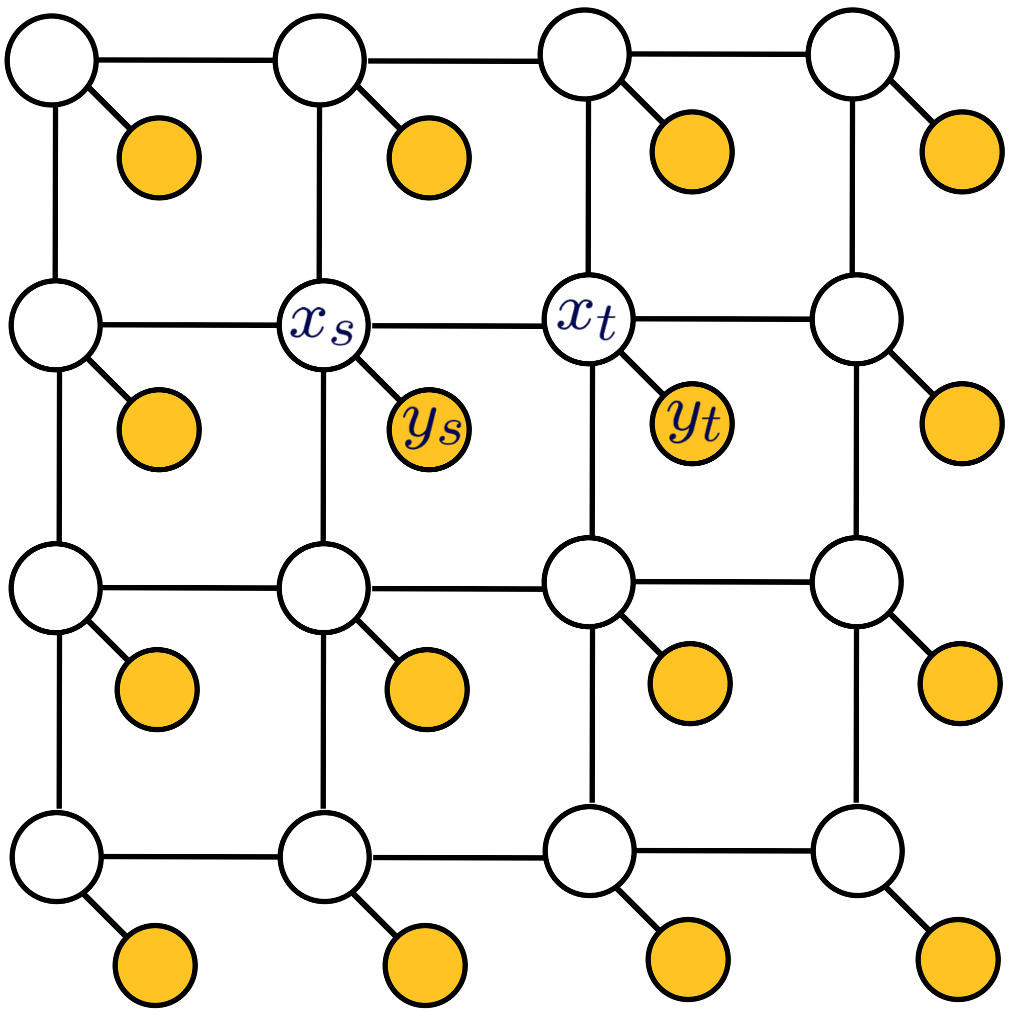
HMM as an Undirected Model



$$p(x, y) = p(x_0) \prod_{t=1}^{T} p(x_t \mid x_{t-1}) p(y_t \mid x_t)$$

"Conditioned on the present, the past and future are statistically independent"

Nearest-Neighbor Grids



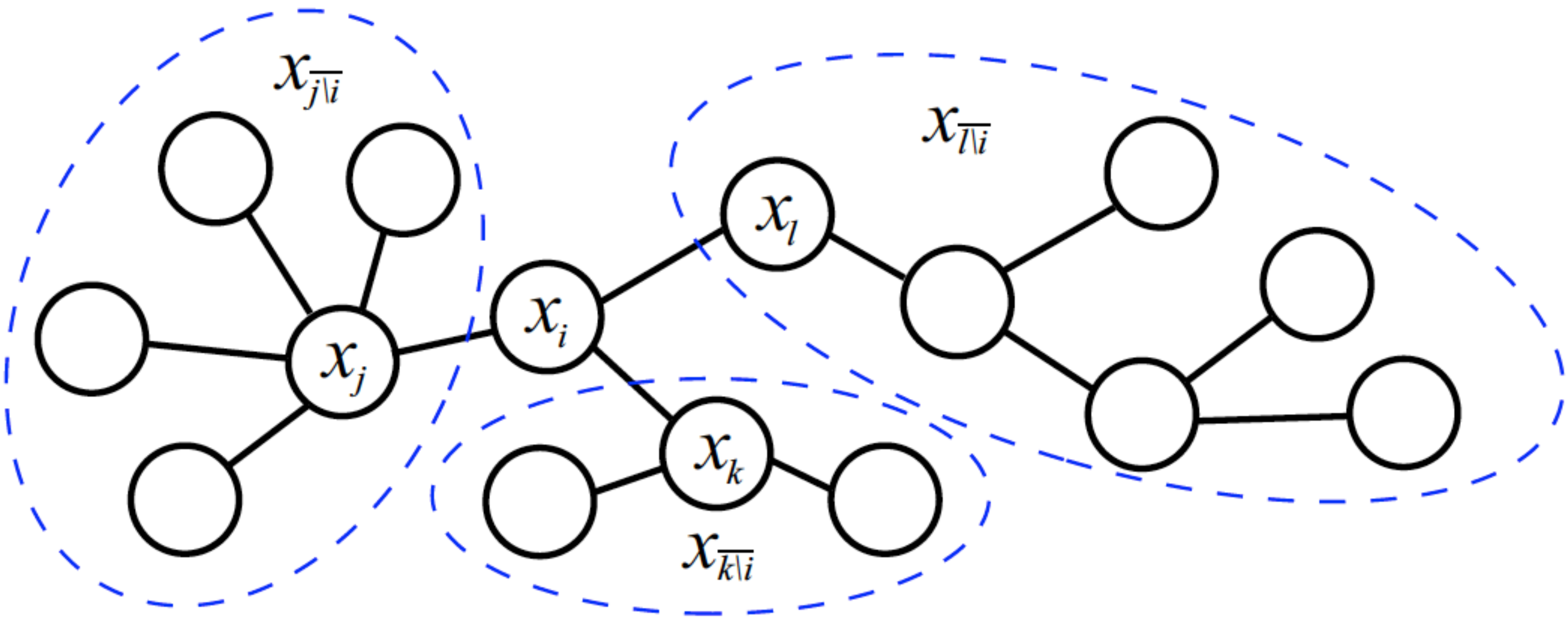
Low Level Vision

- Image denoising
- Stereo
- Optical flow
- Shape from shading
- Superresolution
- Segmentation

x_s → unobserved or hidden variable

y_s → local observation of x_s

Markov Properties in Trees



Directed Conditional Independence

$$p(x_A, x_C \mid x_B) = p(x_A \mid x_B)p(x_C \mid x_B)$$

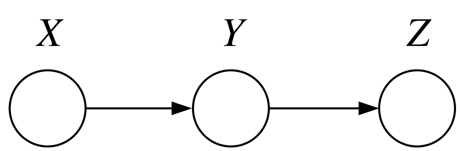
$$p(x_A \mid x_B, x_C) = p(x_A \mid x_B)$$

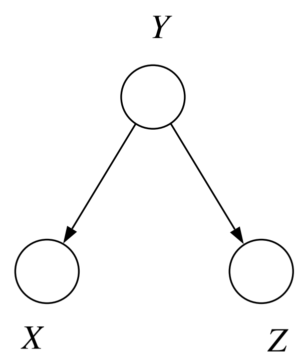
⟺

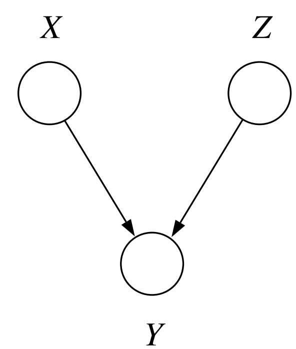
A, C are independent given B

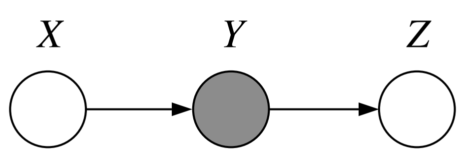
$$A, B, C \subseteq \mathcal{V}$$

GOAL: Characterize conditional independencies which hold for *all* joint distributions which factorize as in a directed graph

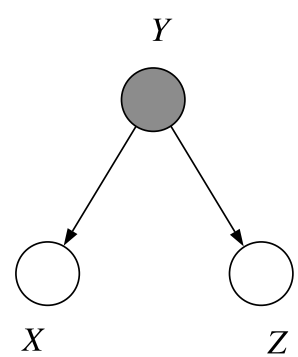




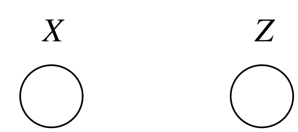




$$X \perp\!\!\!\perp Z \mid Y$$



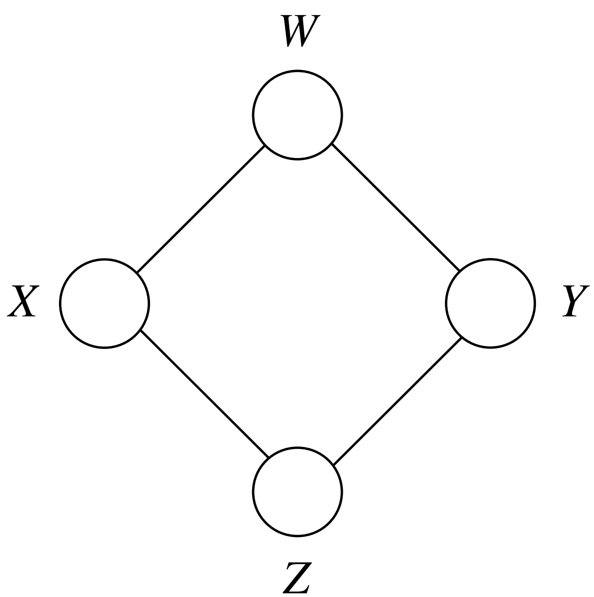
$$X \perp\!\!\!\perp Z \mid Y$$

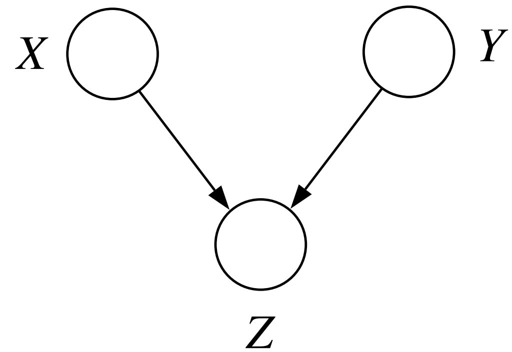


$$X \perp\!\!\!\perp Z$$

Marginally independent
but conditionally dependent!

Markov: Directed vs. Undirected





$$X \perp Y \mid \{W, Z\}$$

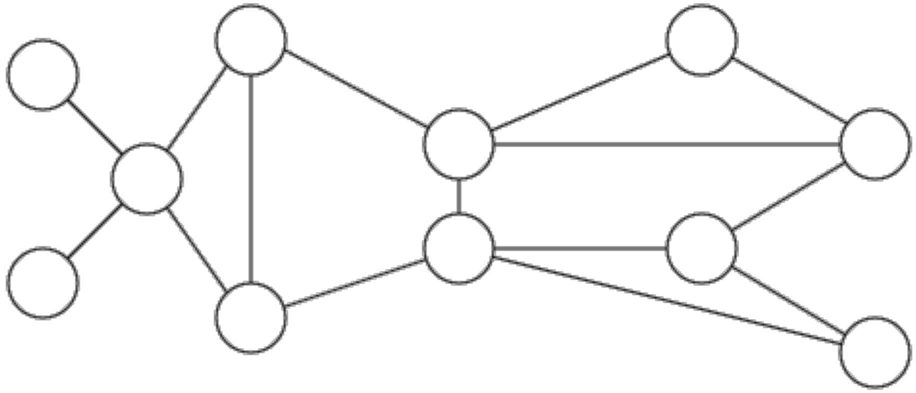
$$W \perp Z \mid \{X, Y\}$$

Can represent one, but not both simultaneously, of these conditional independencies in a single directed model.

$$X \perp Y$$

Graph separation implies that we cannot represent unconditional independence, but conditional dependence, in an undirected model.

Pairwise Markov Random Fields



$$p(x \mid y) = \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s, y)$$

$\mathcal{V}$ $\longrightarrow$ set of N nodes $\{1, 2, \ldots, N\}$

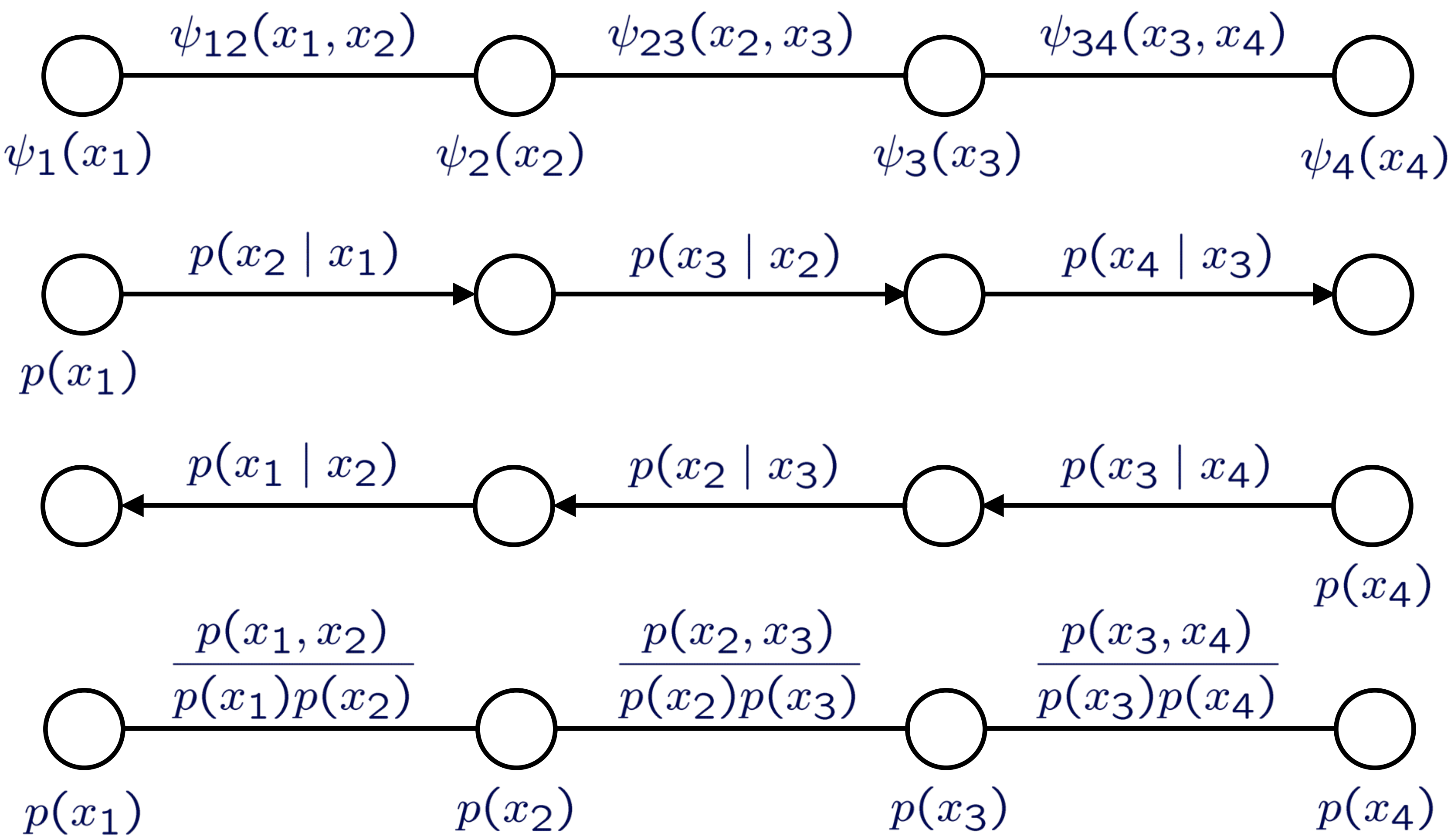
$\mathcal{E}$ $\longrightarrow$ set of edges (s,t) connecting nodes $s, t \in \mathcal{V}$

Z $\longrightarrow$ normalization constant (partition function)

- Product of arbitrary positive *pairwise potential* functions
- Guaranteed Markov with respect to corresponding graph

Markov Chain Factorizations

$$p(x \mid y) = \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s, y)$$



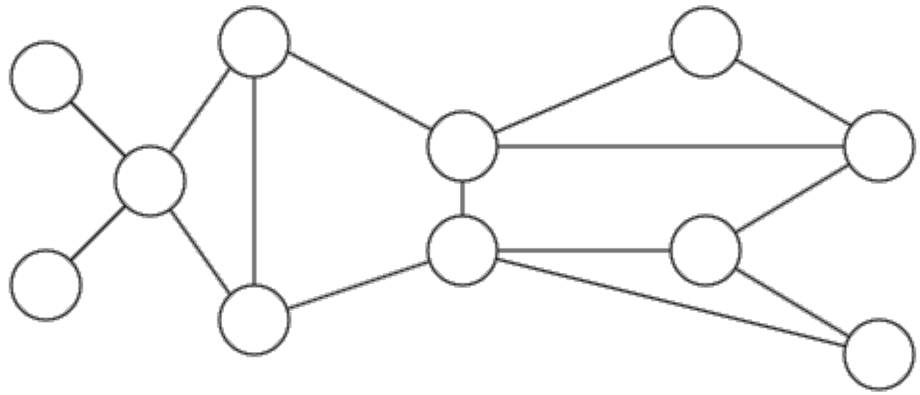
Energy Functions

$$
\begin{aligned}
p(x \mid y) &= \frac{1}{Z} \prod_{(s,t)\in\mathcal{E}} \psi_{st}(x_s, x_t) \prod_{s\in\mathcal{V}} \psi_s(x_s, y) \\
&= \frac{1}{Z} \exp\left\{ - \sum_{(s,t)\in\mathcal{E}} \phi_{st}(x_s, x_t) - \sum_{s\in\mathcal{V}} \phi_s(x_s, y) \right\} \\
&= \frac{1}{Z} \exp\{-E(x)\}
\end{aligned}
$$

$$\phi_{st}(x_s, x_t) = -\log \psi_{st}(x_s, x_t) \qquad \phi_s(x_s) = -\log \psi_s(x_s)$$

- Interpretation inspired by statistical physics
- Justifications from probability (notational convenience)

What Distributions are Markov?



- A *clique* is a fully connected subset of nodes

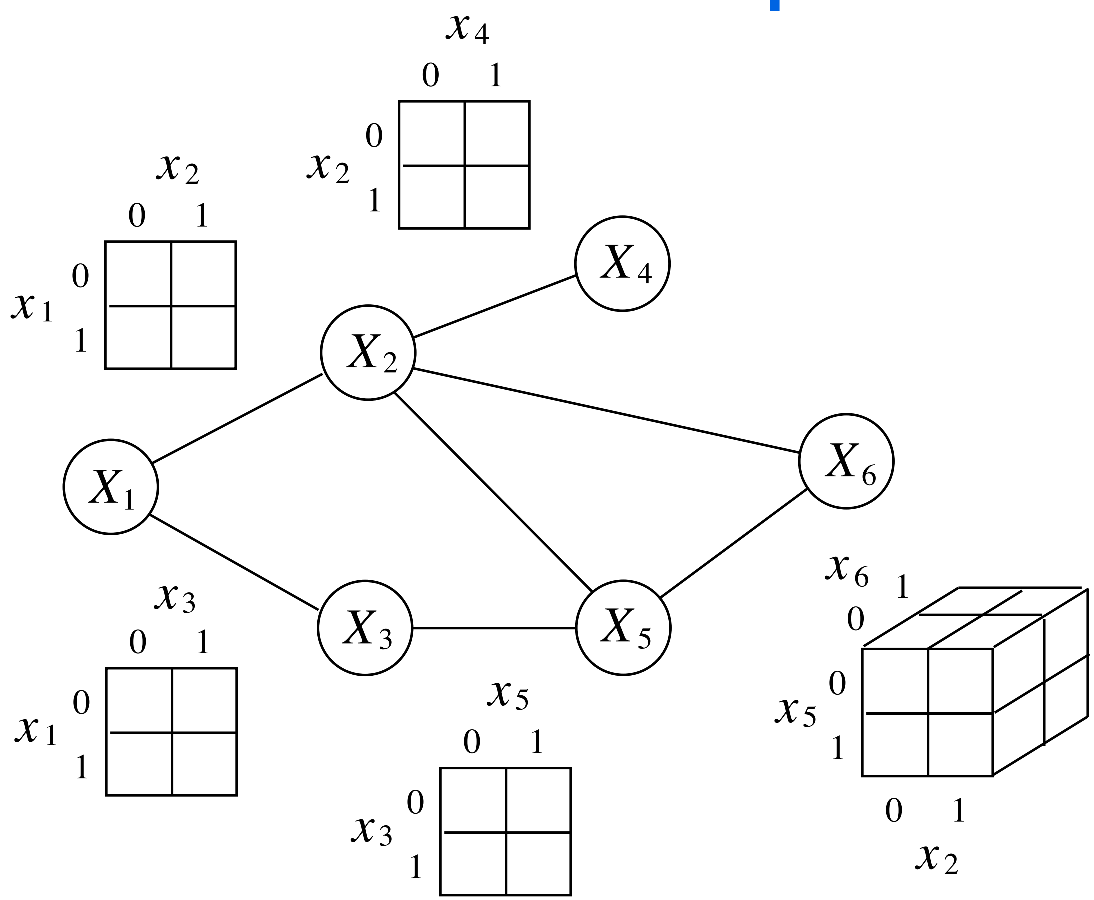
Theorem 2.2.1 (Hammersley-Clifford). *Let $\mathcal{C}$ denote the set of cliques of an undirected graph $\mathcal{G}$. A probability distribution defined as a normalized product of nonnegative potential functions on those cliques is then always Markov with respect to $\mathcal{G}$:*

$$p(x) \propto \prod_{c \in \mathcal{C}} \psi_c(x_c) \tag{2.71}$$

Conversely, any strictly positive density ($p(x) > 0$ for all x) which is Markov with respect to $\mathcal{G}$ can be represented in this factored form.

- It is possible, but not necessary, to restrict factorization only to the *maximal cliques* (not strict subsets of other cliques)

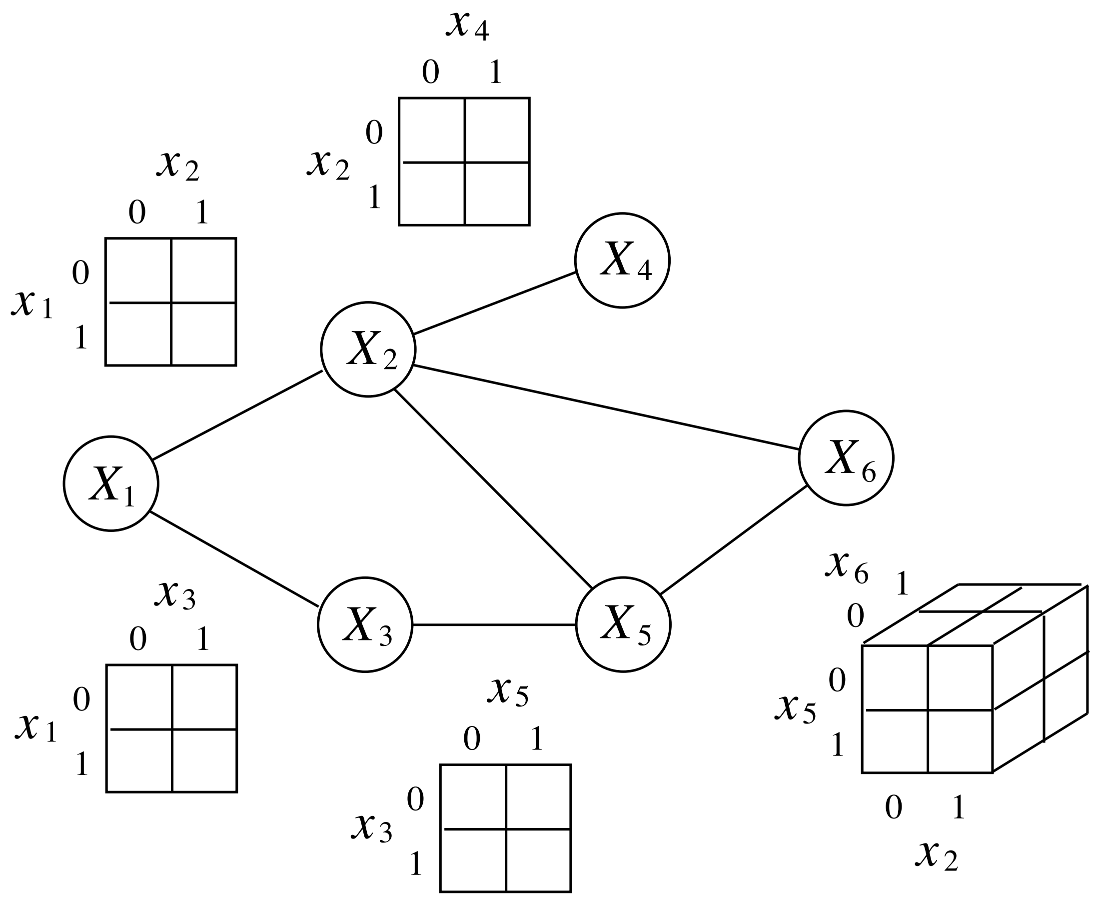
Parameterization & Representation



Representational (storage, learning, computation) Complexity

- *Joint distribution:* Exponential in number of variables
- *Undirected graphical model:* Exponential in number of variables contained in the *maximal cliques* of the graph

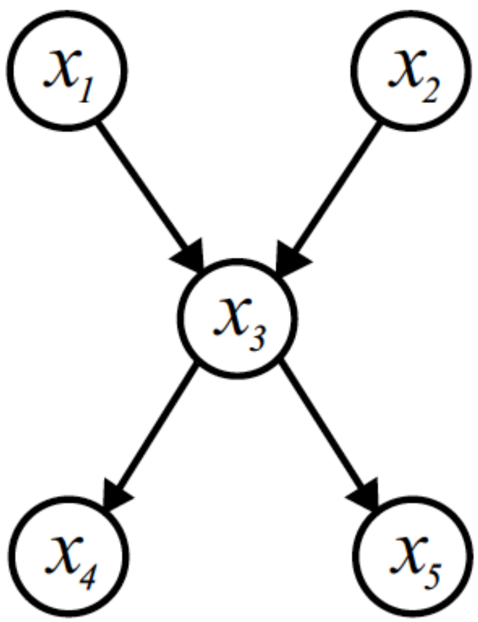
Potential Confusions



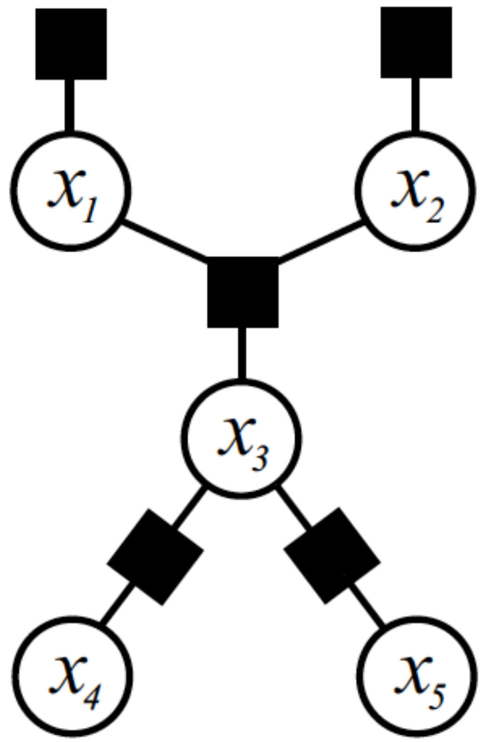
For graphs with cycles:

- *Potential functions* usually are not marginal probabilities
- *Conditional distributions* of nodes given neighbors cannot be independently specified, and guarantee a valid joint

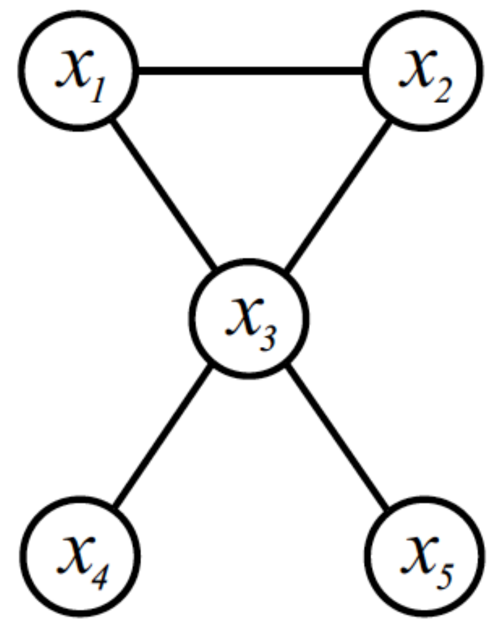
Types of Graphical Models



Directed



Factor



Undirected