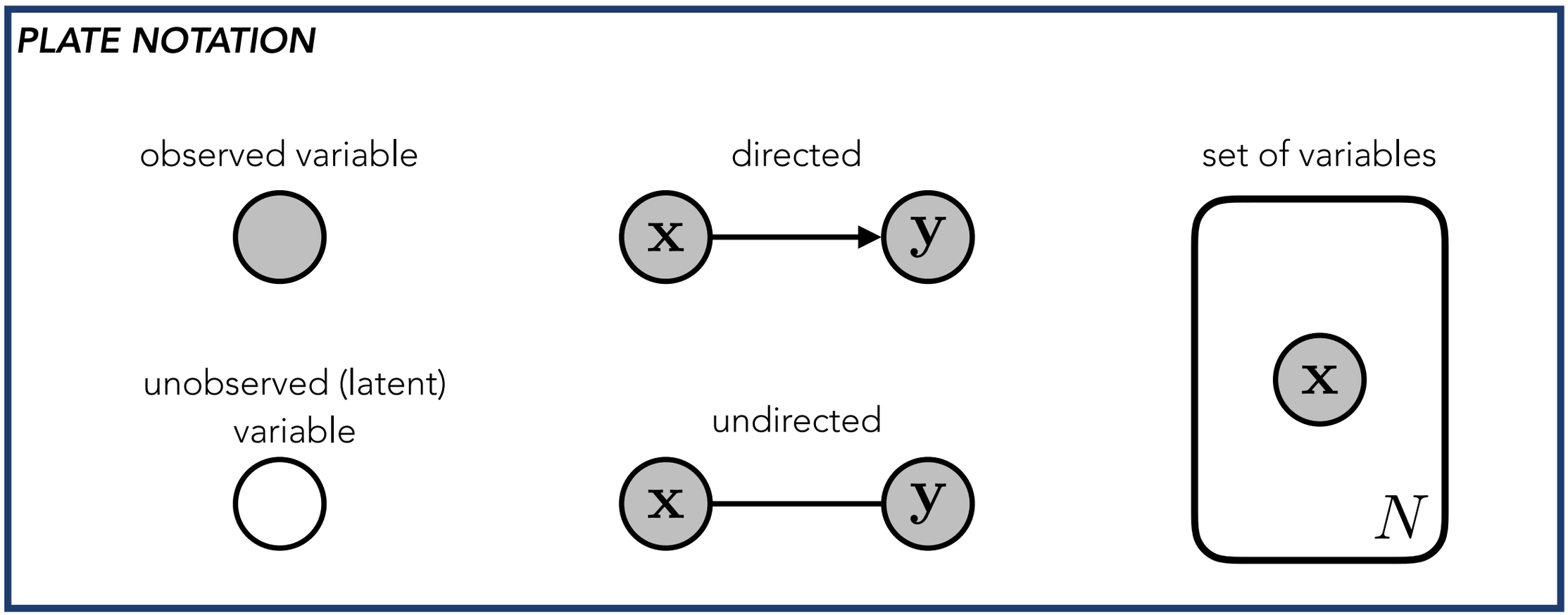
Northeastern

CS 7140:
ADVANCED MACHINE LEARNING

GRAPHICAL MODEL

probabilistic graphical models provide a framework
for modeling relationships between random variables



BAYESIAN NETWORK

Bayesian Network

- Directed Acyclic Graph (DAG)
  - One node for each random variable X_i
  - One conditional distribution per node $P(X_i \mid Pa(X_i))$

- Chain rule:

$$
\begin{aligned}
P(X_4, X_3, X_2, X_1) &= P(X_4 \mid X_3, X_2, X_1) \cdot P(X_3, X_2, X_1) \\
&= P(X_4 \mid X_3, X_2, X_1) \cdot P(X_3 \mid X_2, X_1) \cdot P(X_2, X_1) \\
&= P(X_4 \mid X_3, X_2, X_1) \cdot P(X_3 \mid X_2, X_1) \cdot P(X_2 \mid X_1) \cdot \mathrm{P}(X_1)
\end{aligned}
$$

$$
P(X_1, X_2, \cdots, X_n) = \prod_i P(X_i \mid X_1, \cdots, X_{i-1})
$$

- Bayesian Network chain rule:

$$
P(X_1, X_2, \cdots, X_n) = \prod_i P(X_i \mid Pa(X_i))
$$

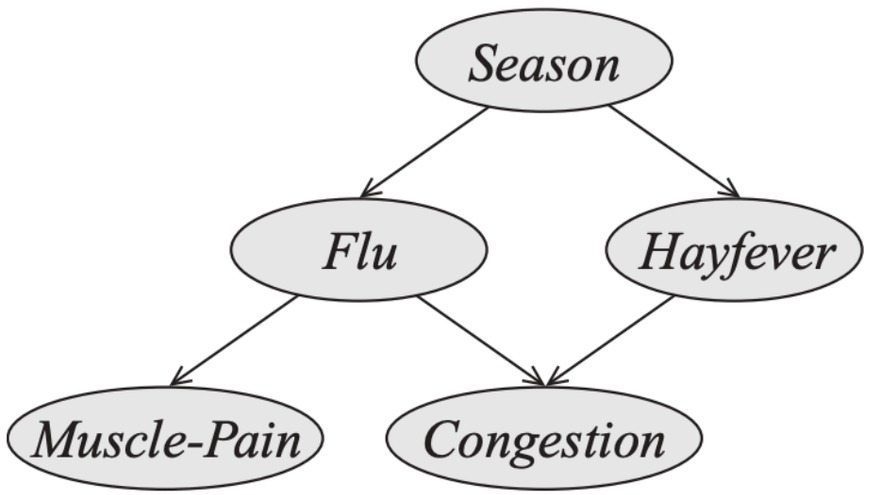
Bayesian Network

- Chain rule:

$$P(S,F,H,C,M) = P(S)P(F\,|\,S)P(H\,|\,S,F)P(C\,|\,H,F,S)P(M\,|\,C,H,F,S)$$

- Bayesian Network chain rule

$$P(S,F,H,C,M) = P(S)P(F\,|\,S)P(H\,|\,S)P(C\,|\,H,F)P(M\,|\,F)$$



- Conditional Independence

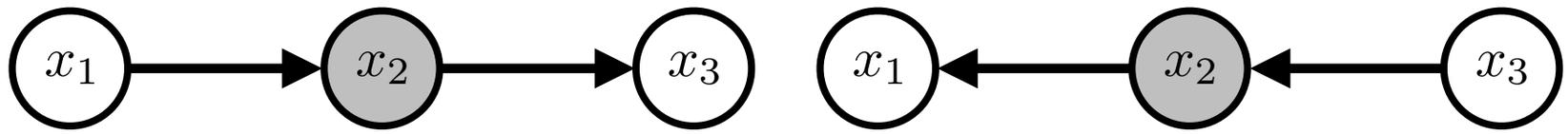
$$H \perp F\,|\,S, \quad C \perp S\,|\,F,H, \quad M \perp C,H,S\,|\,F$$

- Bayesian Network encodes conditional independence

D(dependence)-Separation

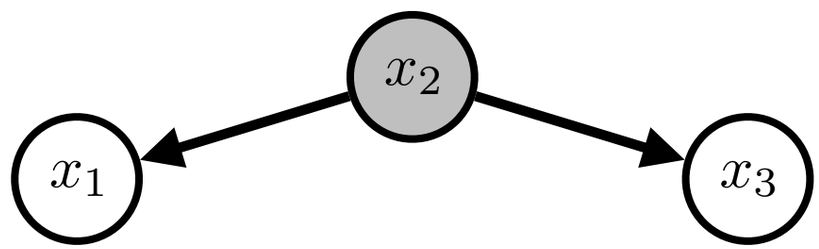
- Little Sequence: causal/evidence trail

$X_1 \perp X_3 | X_2$



- Little Tree: common cause

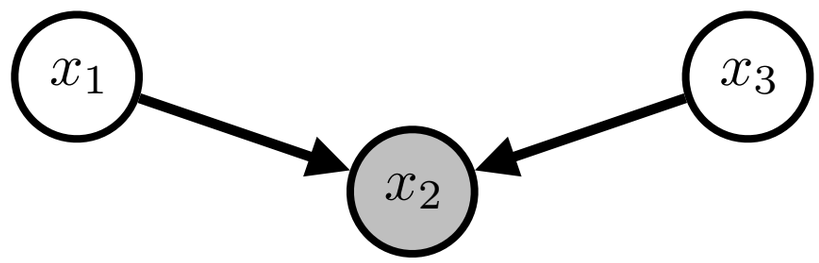
$X_1 \perp X_3 | X_2$



- Little V: common effect

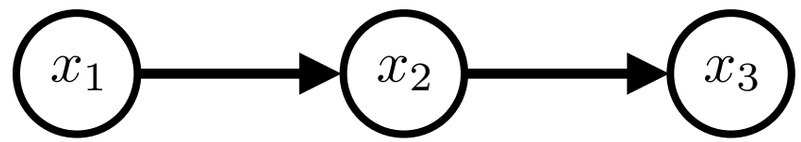
$X_1 \not\perp X_3 | X_2$

explaining away

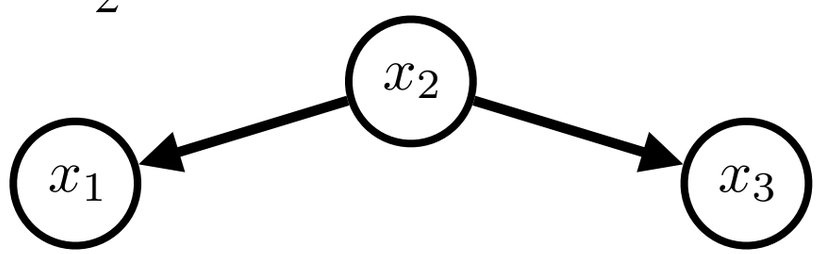


D(dependence)-Separation

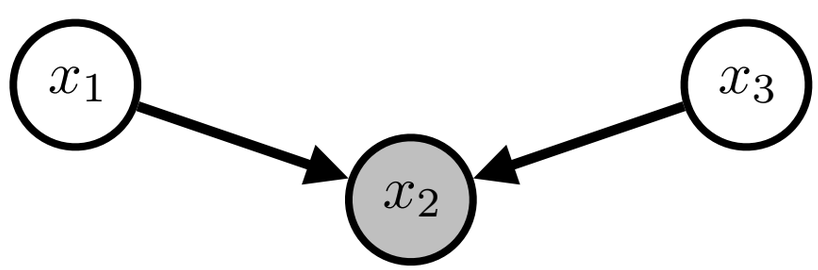
- Active Trail: when influence can flow
  - **causal/evidence trail**: active if X_2 is latent



  - **common cause**: active if X_2 is latent



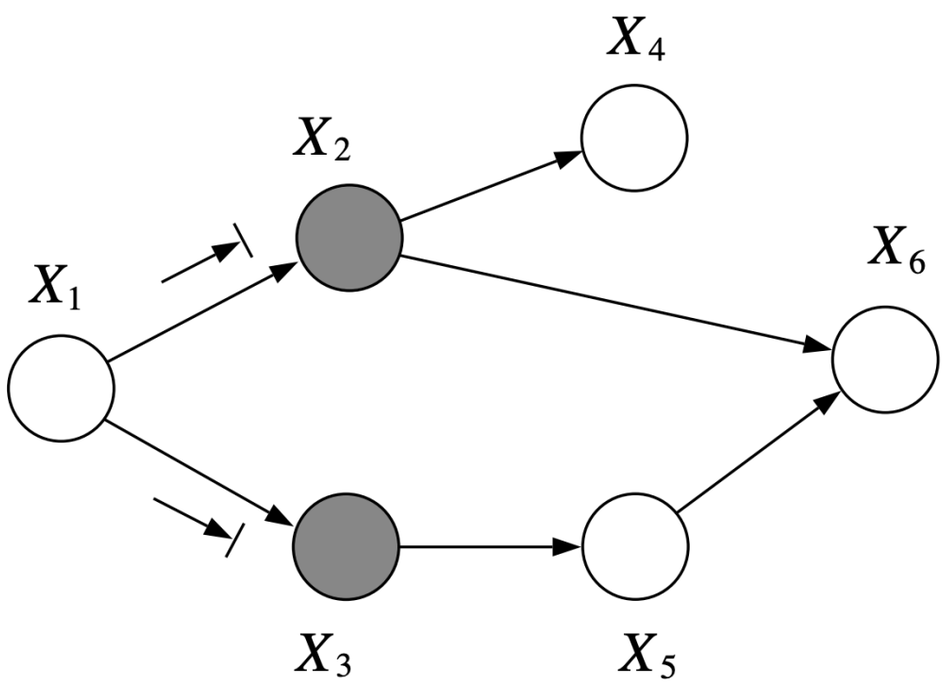
  - **common effect**: if active if X_2 or X_2 descendant is observed

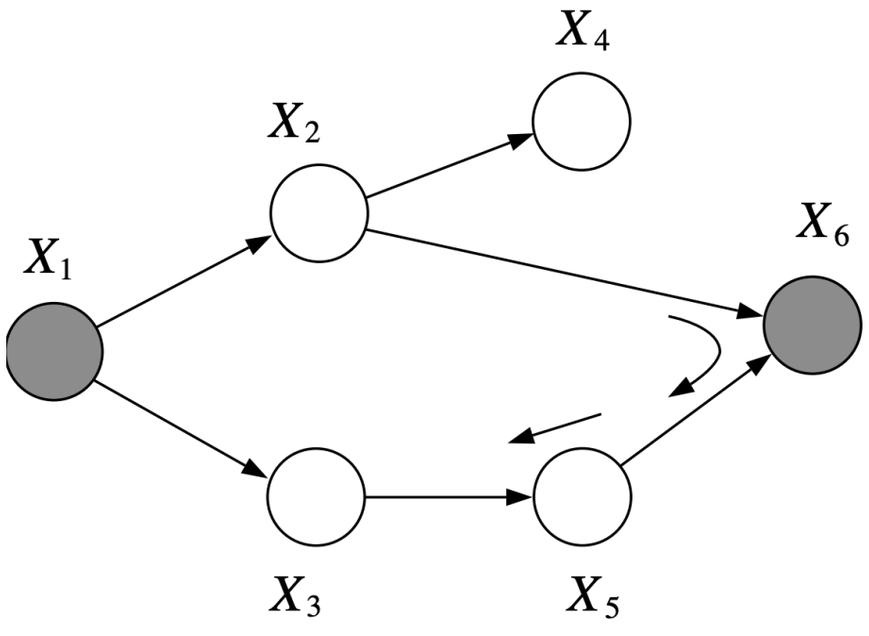


D(dependence)-Separation

- Check **active trail** between X and Z when variables Y are observed

- d-separation reduces statistical independencies (hard) to connectivity in graphs (easy)

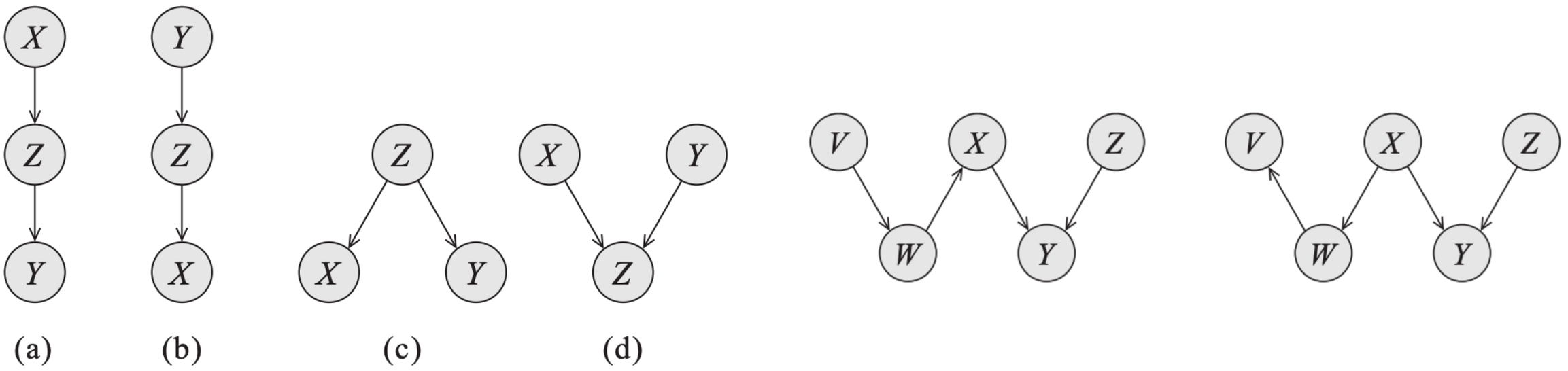




I-Equivalence

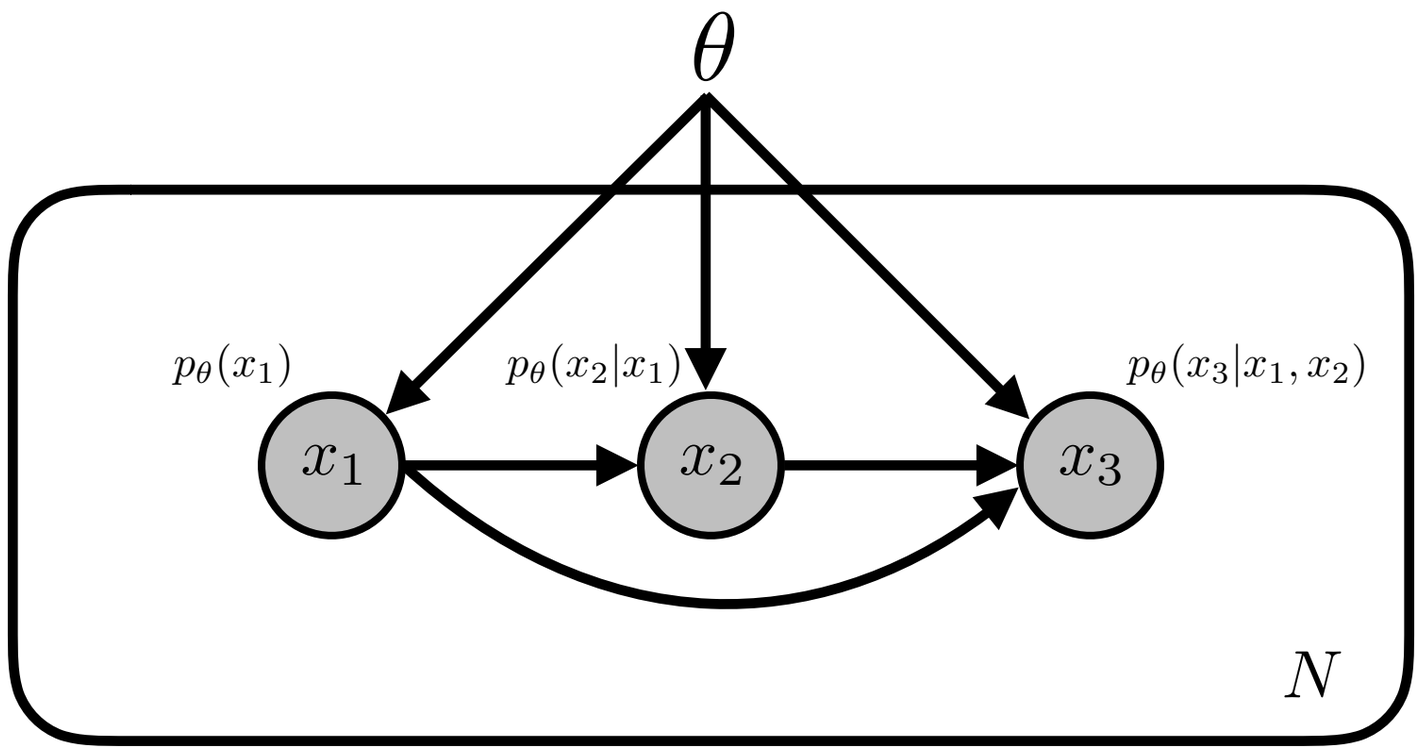
- Two Bayesian networks are I-equivalent if they encode precisely the same conditional independence assertions.

- Are these I-equivalent?

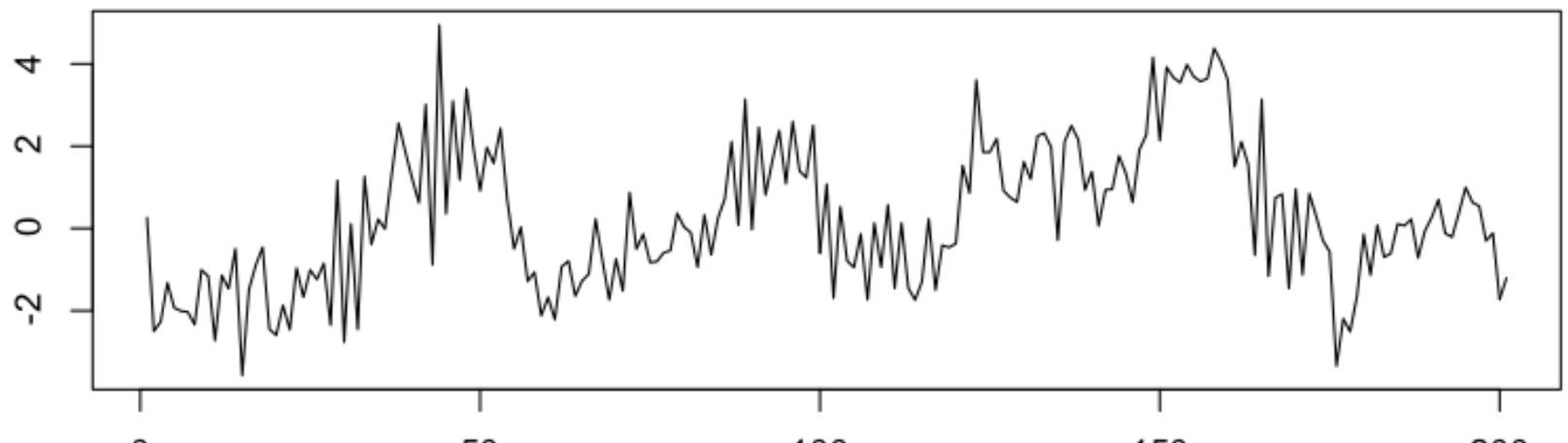


Two Bayesian networks are I-equivalent if they have the same **skeletons** and same **V-structures**.

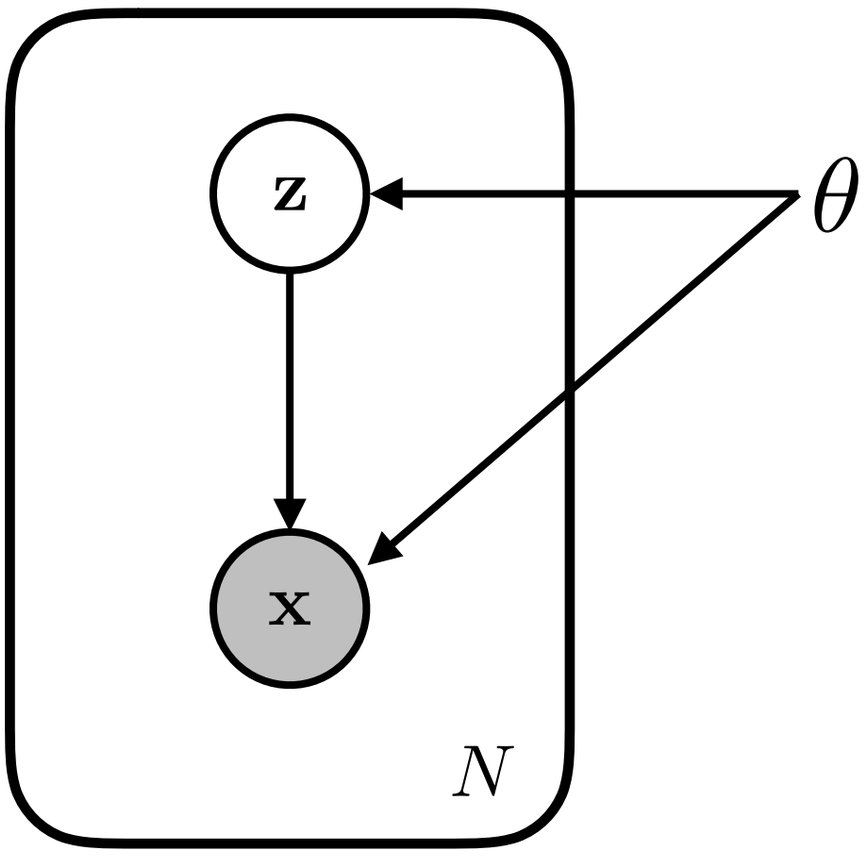
Autoregressive model



$$p(x_1, x_2, x_3) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2)$$



Latent Variable Model

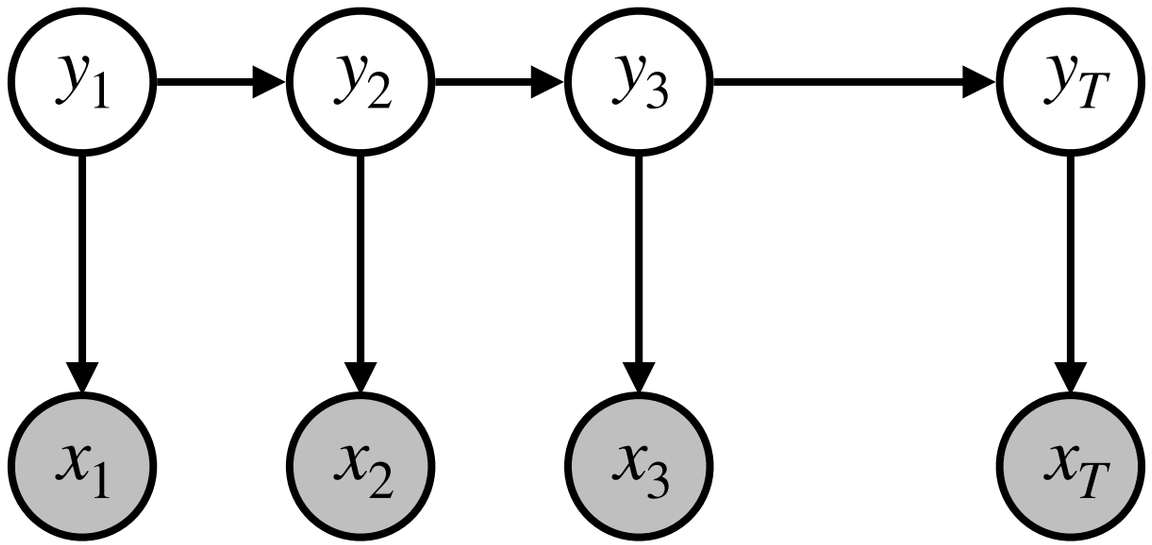


$$p(x, z) = p(z)p(x \mid z)$$

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

Hidden Markov Model



$$p(x_1, \cdots, x_T, y_1, \cdots, y_T) = p(y_1)p(x_1 \mid y_1) \prod_{t=2}^{T} p(y_t \mid y_{t-1})p(x_t \mid y_t)$$

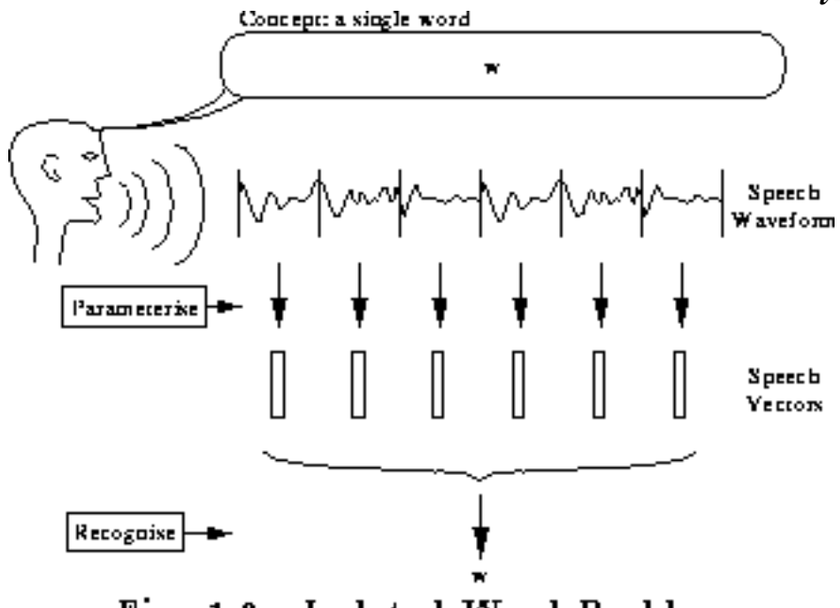
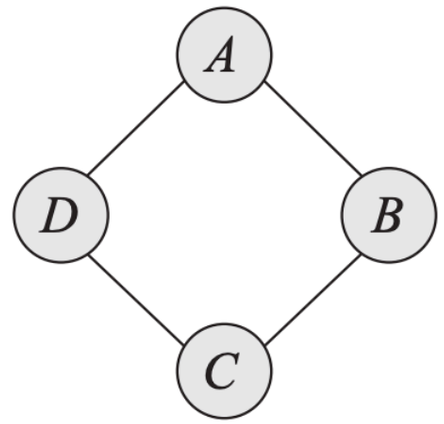


Fig. 1.2 Isolated Word Problem

MARKOV RANDOM FIELD

Markov Random Field

- Undirected graph
  - One node for each random variable



- Positive density
- Satisfy **Markov property**:
  - pairwise: $X_u \perp X_v \mid X_{V\setminus\{u,v\}}$
  - local: $X_v \perp X_{V\setminus N(v)} \mid X_{N(v)}$
  - global: $X_A \perp X_b \mid X_S$

$$(A \perp C \mid B, D)$$
$$(B \perp D \mid A, C)$$

$$P(A,B,C,D) = \tfrac{1}{Z}\phi_1(A,B)$$
$$\phi_2(B,C)\phi_3(C,D)\phi_4(A,D)$$

Markov Random Field

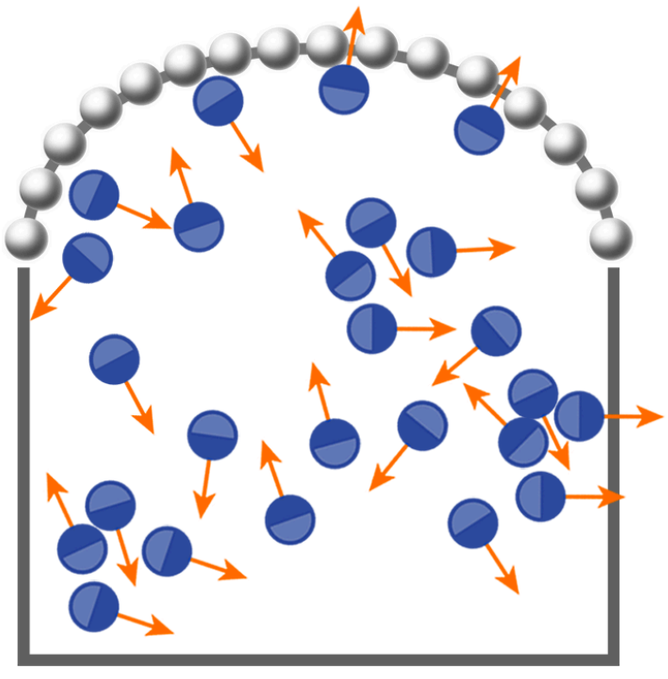
- **Boltzmann-Gibbs Distribution**:

  - Potential function over cliques of the graph

$$p(X_1, X_2, \cdots, X_T) = \frac{1}{Z}\prod_{c \in C} \phi_c(X_c)$$

  - Z is the partition function

$$Z = \sum_{X_1, X_2, \cdots, X_n} \prod_{c \in C} \phi_c(X_c)$$



- The factors do not correspond either to probabilities or to conditional probabilities.

$$P = \frac{1}{Z} e^{-\frac{\epsilon}{kT}}$$

Hammersley–Clifford theorem

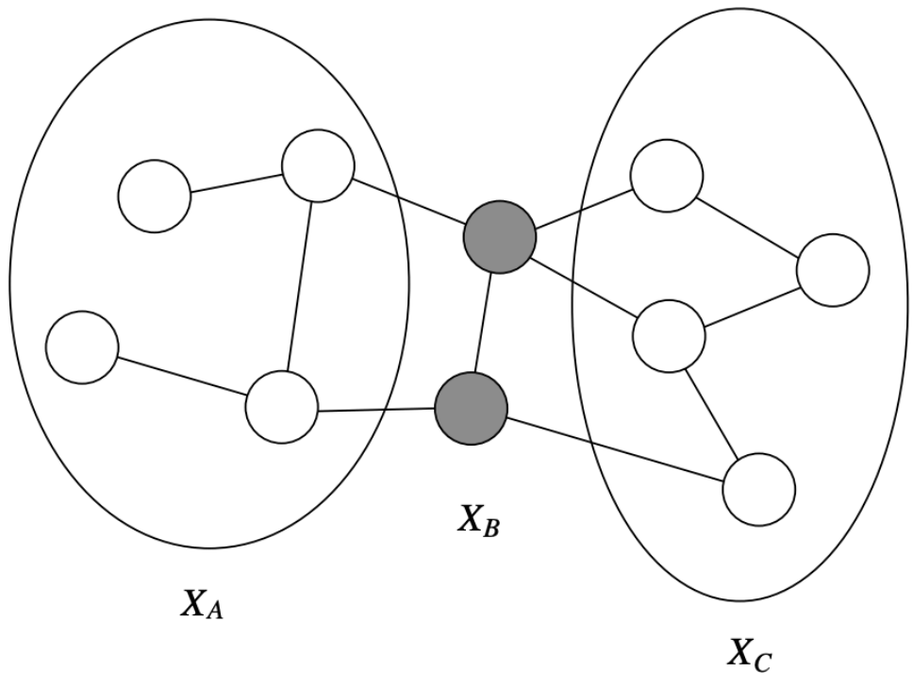
A probability distribution that has a strictly positive density satisfies one of the Markov properties with respect to an undirected graph *G* if and only if it is a Gibbs distribution, that is, its density can be factorized over the cliques of the graph.

- fundamental theorem of random fields
- equivalence of Markov random field and Gibbs distribution
- conditional independence vs factorization

Markov Network Independence

- Conditional independence is given by graph separation!

- $X_A \perp X_C | X_B$ if there is no path from a ∈ A to c ∈ C after removing all variables in B

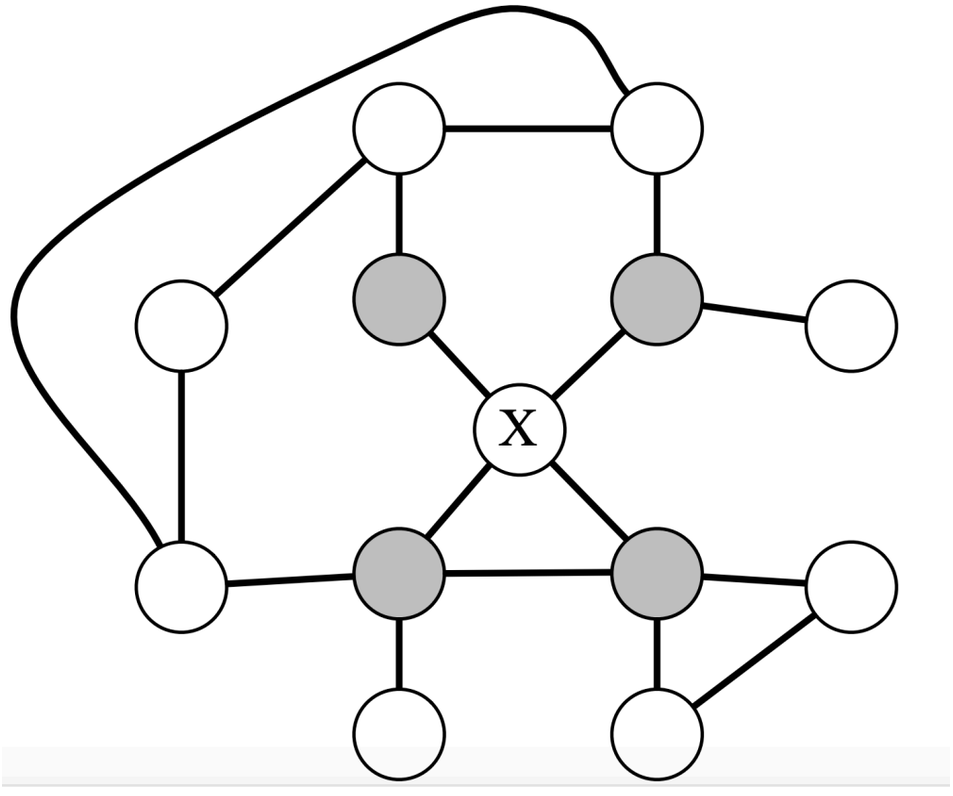


Markov Blanket

- A set U is a Markov blanket of $\mathcal{X}$ in a distribution P if $X \notin U$ and if U is a minimal set of Markov blanket nodes such that

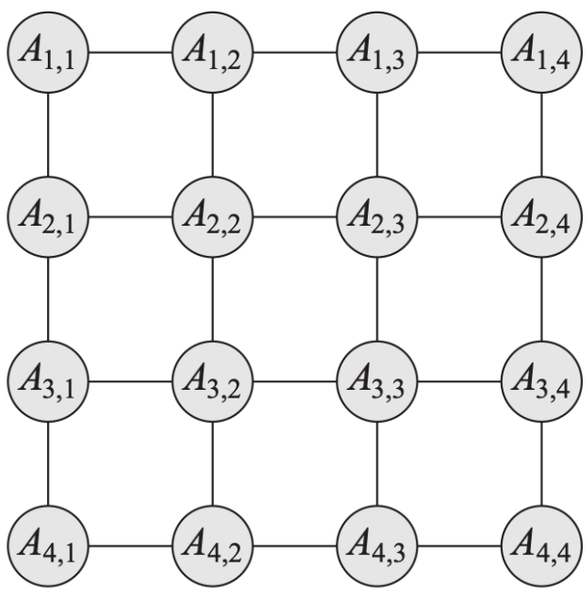
$$(X \perp \mathcal{X} - \{X\} - U \mid U) \in I(P)$$

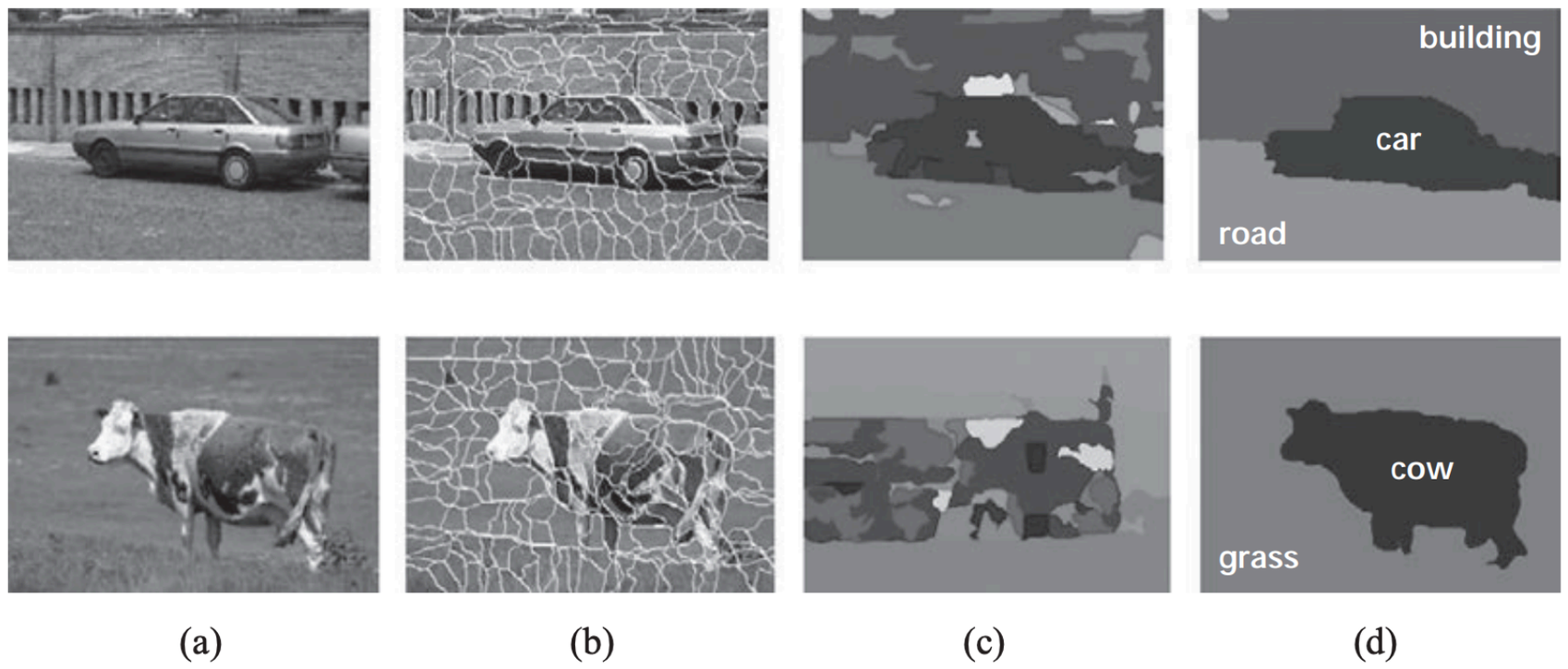
- In undirected graphical models, the Markov blanket of a variable is precisely its neighbors in the graph:



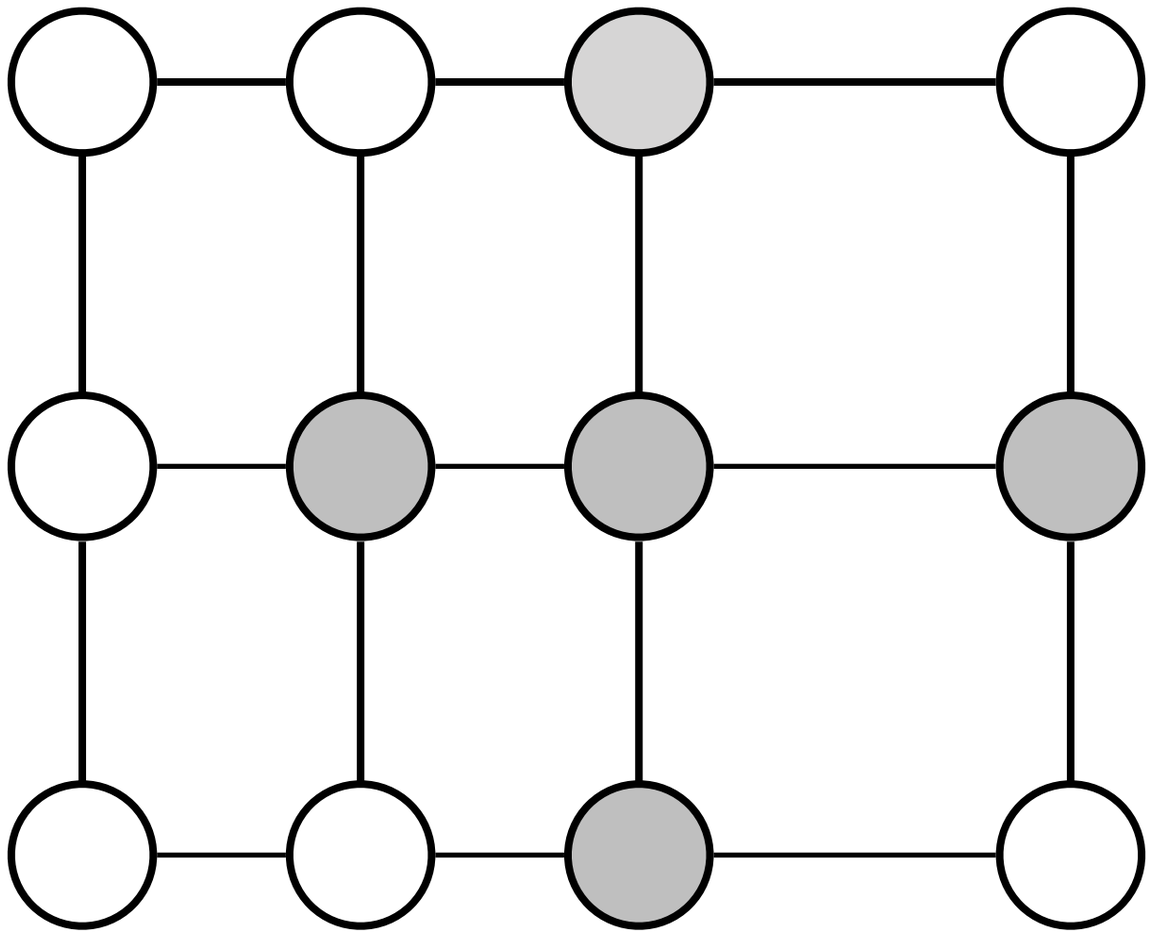
Example Application

- image segmentation: original image, superpixels, node potential alone, edge potential
- The importance of modeling the correlations between the superpixels.





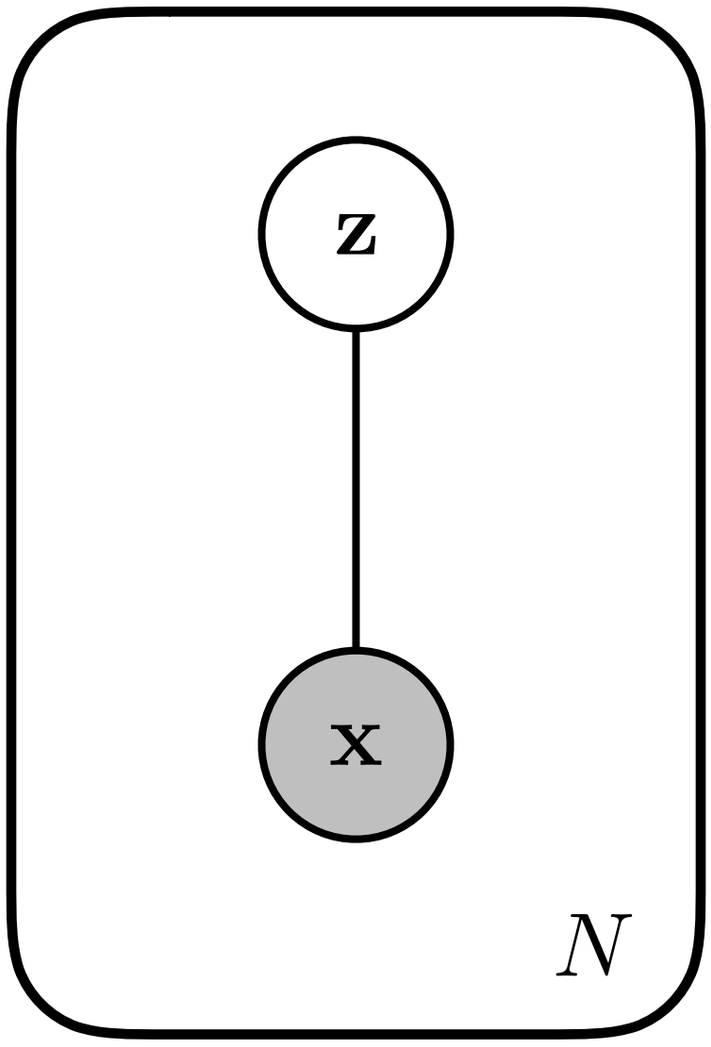
Ising Model



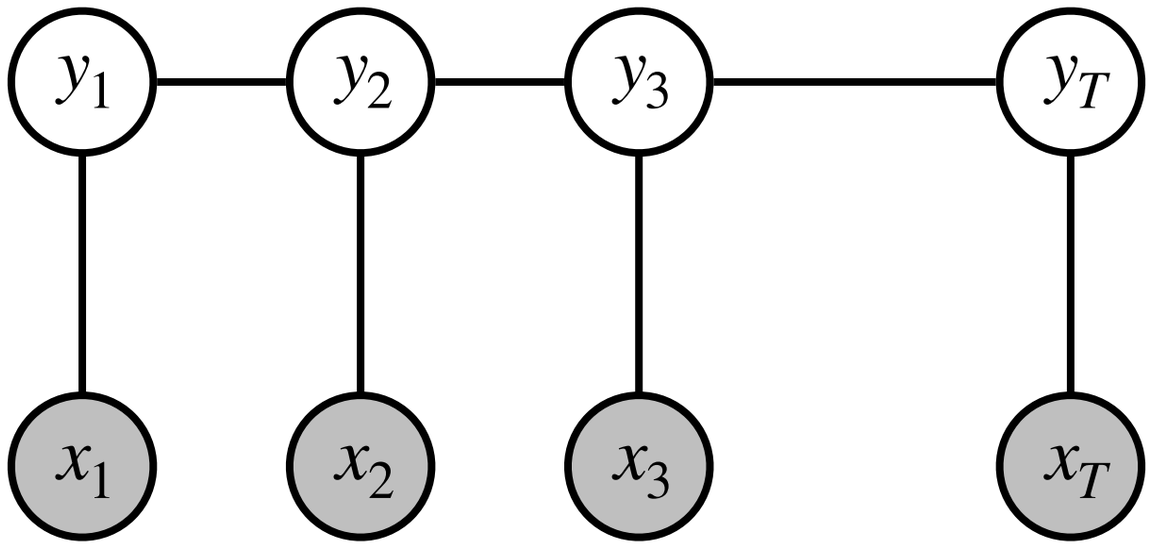
$$p(x_1, \cdots, x_T) = \frac{1}{Z} \prod_{i,j} \phi(x_i, x_j)$$

Latent Variable Model

→ restricted Boltzmann machine



Conditional Random Field



$$p(y_1, \cdots, y_T | x_1, \cdots, x_T) = \frac{1}{Z} \prod_{t-1}^{T-1} \phi(y_t, y_{t-1}) \prod_{t=1}^{T} \phi(x_t, y_t)$$