Computer
Science

CSC535: Probabilistic Graphical Models

Probabilistic Graphical Models

Prof. Jason Pacheco

Administrivia

- HW1 will be graded by end of week
- HW2 will be out this Wednesday

Outline

Directed graphical models

- Bayes Nets
- Conditional dependence

Undirected graphical models

- Markov random fields (MRFs)
- Factor graphs

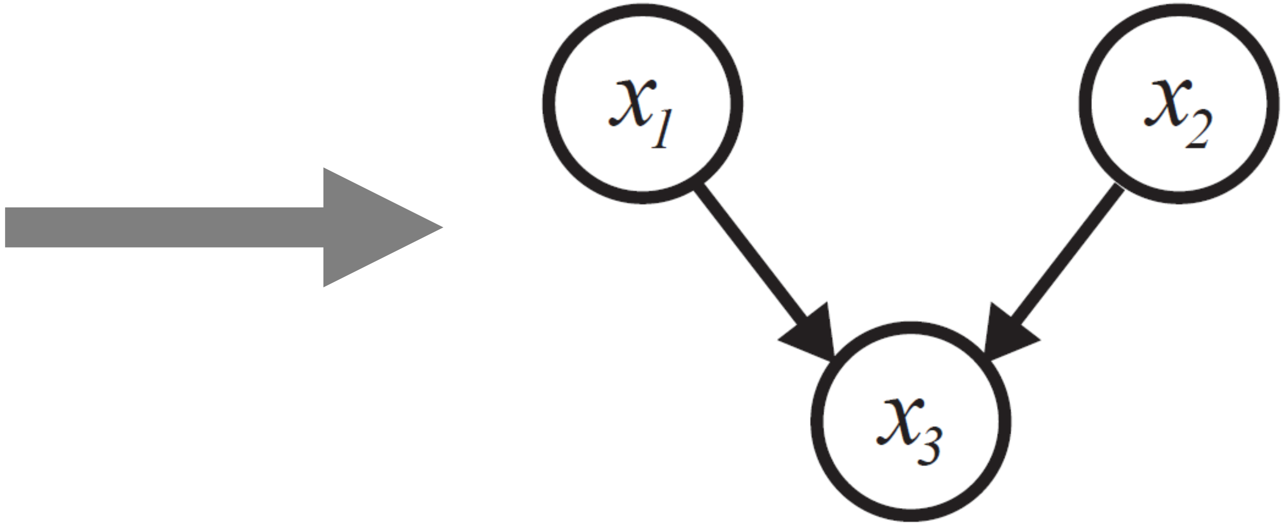
From Probabilities to Pictures

*A probabilistic graphical model allows us to pictorially represent a probability distribution**

Probability Model:

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

Graphical Model:

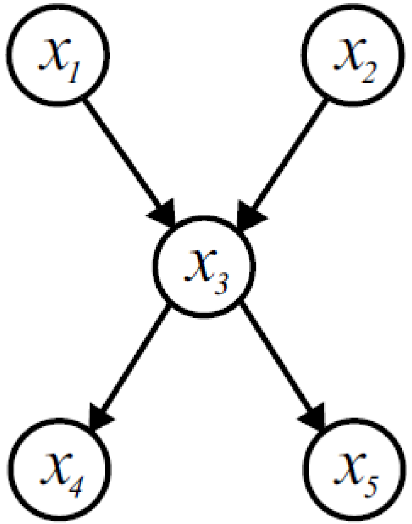


The graphical model structure *obeys* the factorization of the probability function in a sense we will formalize later

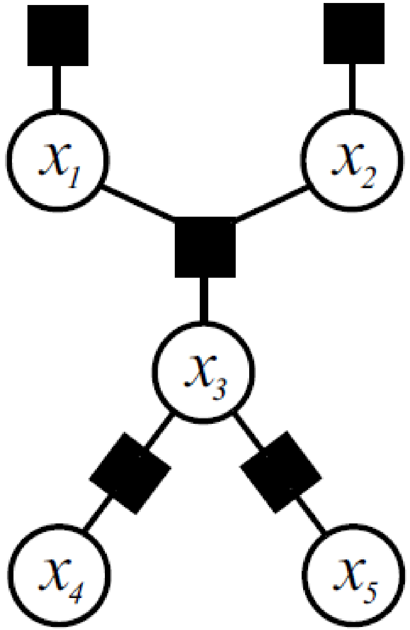
* We will use the term "distribution" loosely to refer to a CDF / PDF / PMF

Graphical Models

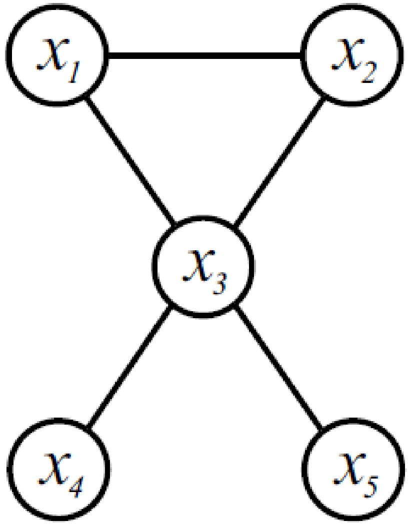
A variety of graphical models can represent the same probability distribution



Bayes Network



Factor Graph



Markov Random Field

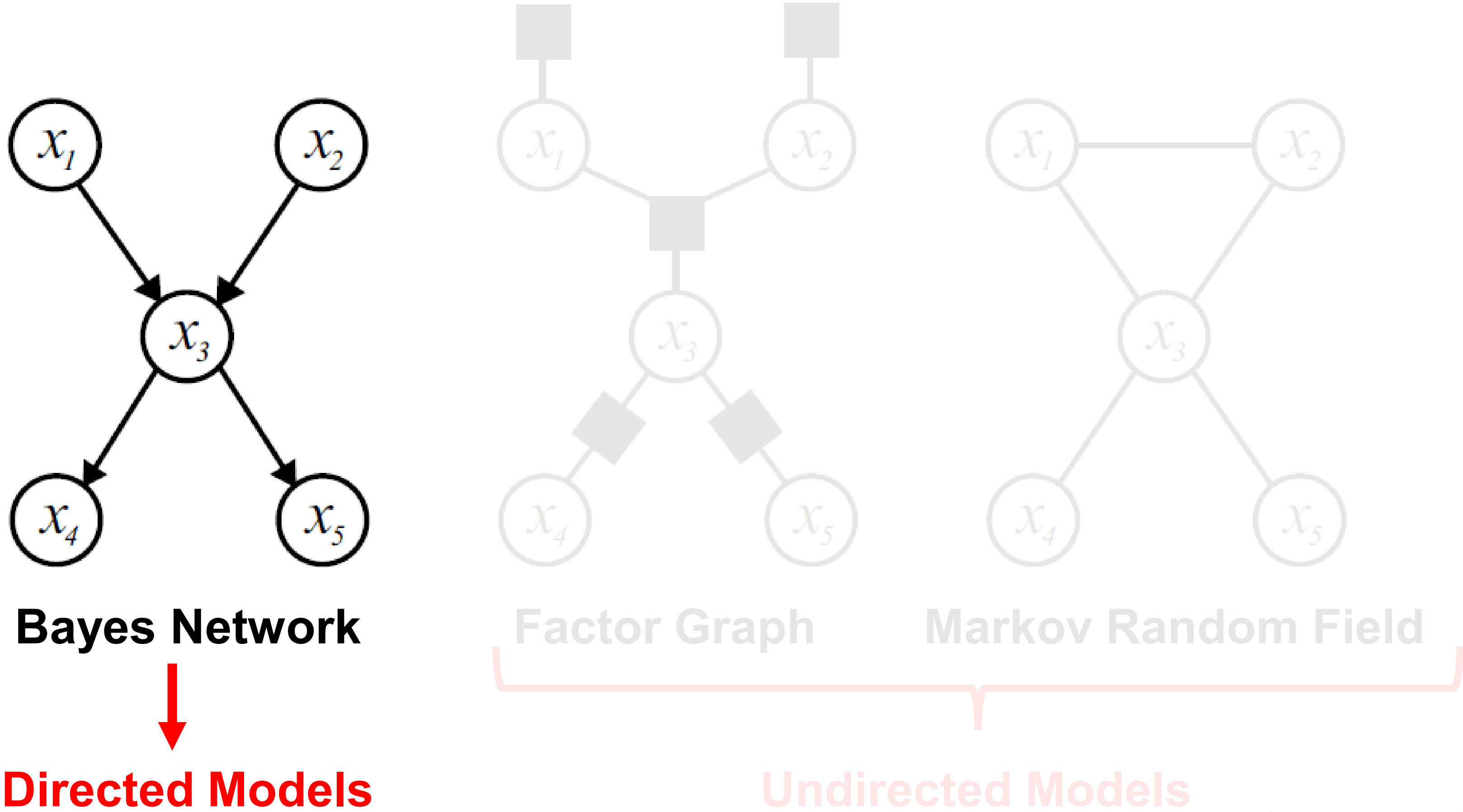
Directed Models

Undirected Models

[Source: Erik Sudderth, PhD Thesis]

Graphical Models

A variety of graphical models can represent the same probability distribution



Bayes Network

Factor Graph

Markov Random Field

Directed Models

Undirected Models

[Source: Erik Sudderth, PhD Thesis]

Chain Rule of Probability

Recall the **probability chain rule** says that we can decompose any joint distribution as a product of conditionals….

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

Valid for *any ordering* of the random variables…

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

For a collection of N RVs and any permutation ρ:

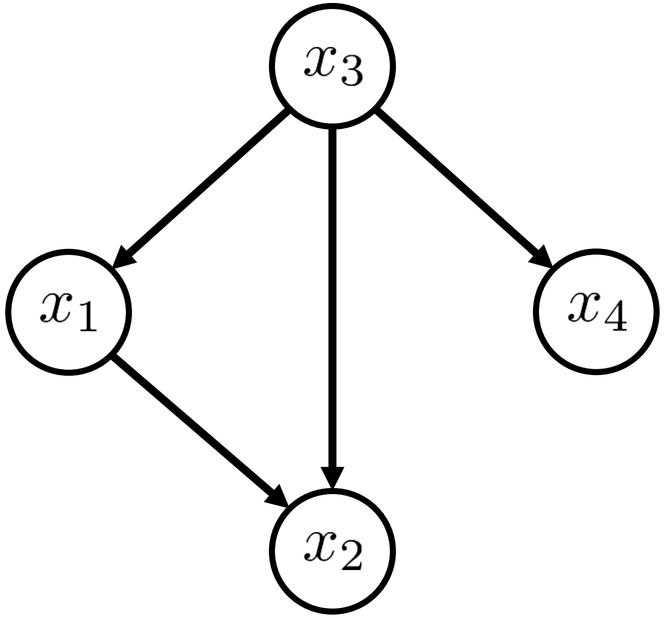
$$p(x_1, \ldots, x_N) = p(x_{\rho(1)}) \prod_{i=2}^{N} p(x_{\rho(i)} \mid x_{\rho(i-1)}, \ldots, x_{\rho(1)})$$

Conditional Independence

Recall two RVs X and Y are **conditionally independent** given Z (or $X \perp Y \mid Z$) iff:

$$p(X \mid Y, Z) = p(X \mid Z)$$

Idea Apply *chain rule* with ordering that exploits conditional independencies to simplify the terms



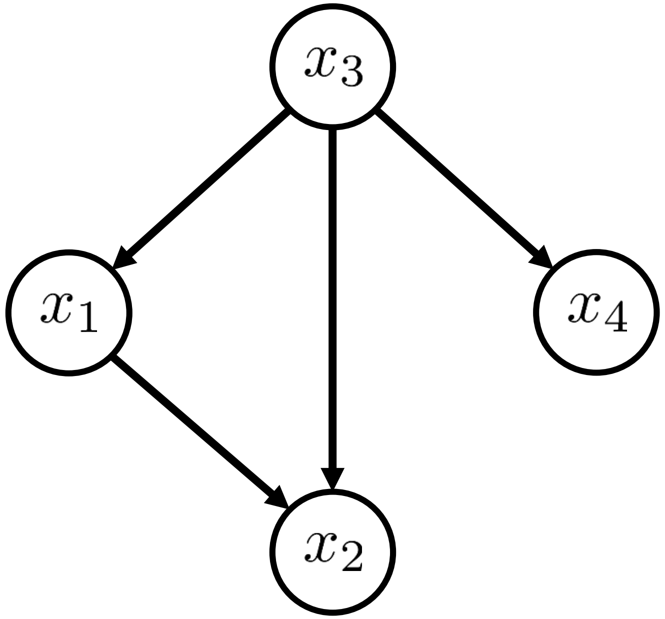
Ex. Suppose $x_4 \perp x_1 \mid x_3$ and $x_2 \perp x_4 \mid x_1$ then:

$$\begin{aligned} p(x) &= p(x_3)p(x_1 \mid x_3)p(x_4 \mid x_1, x_3)p(x_2 \mid x_1, x_3, x_4) \\ &= p(x_3)p(x_1 \mid x_3)p(x_4 \mid x_3)p(x_2 \mid x_1, x_3) \end{aligned}$$

Can visualize conditional dependencies using ***directed acyclic graph*** *(DAG)*

Directed Graphs

Def. A <u>directed graph</u> is a graph with edges $(s, t) \in \mathcal{E}$ (arcs) connecting parent vertex $s \in \mathcal{V}$ to a child vertex $t \in \mathcal{V}$



Def. <u>Parents</u> of vertex $t \in \mathcal{V}$ are given by the set of nodes with arcs pointing to t,

$$\text{Pa}(t) = \{s : (s, t) \in \mathcal{E}\}$$

<u>Children</u> of $t \in \mathcal{V}$ are given by the set,

$$\text{Ch}(t) = \{t : (t, k) \in \mathcal{E}\}$$

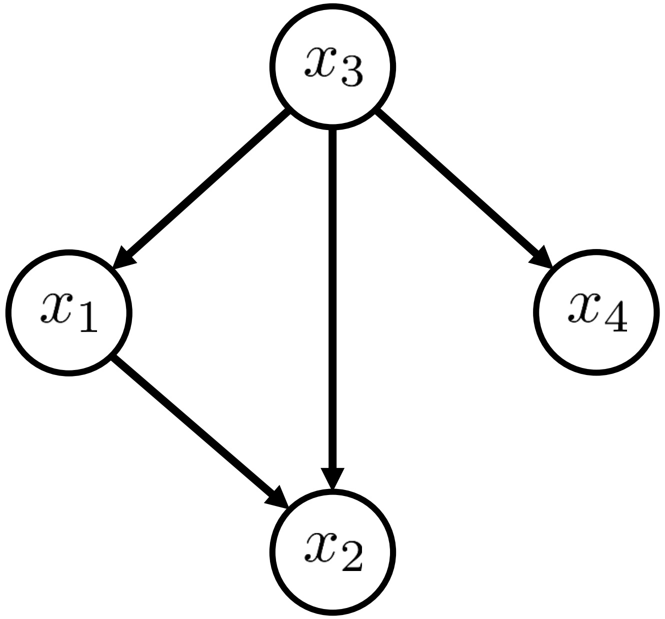
<u>Ancestors</u> are parents-of-parents.
<u>Descendants</u> are children-of-children.

Bayes Network

Model factors are normalized conditional distributions:

$$p(x) = \prod_{s \in \mathcal{V}} p(x_s \mid x_{\mathrm{Pa}(s)})$$

Parents of node *s*



Directed acyclic graph (DAG) specifies factorized form of joint probability:

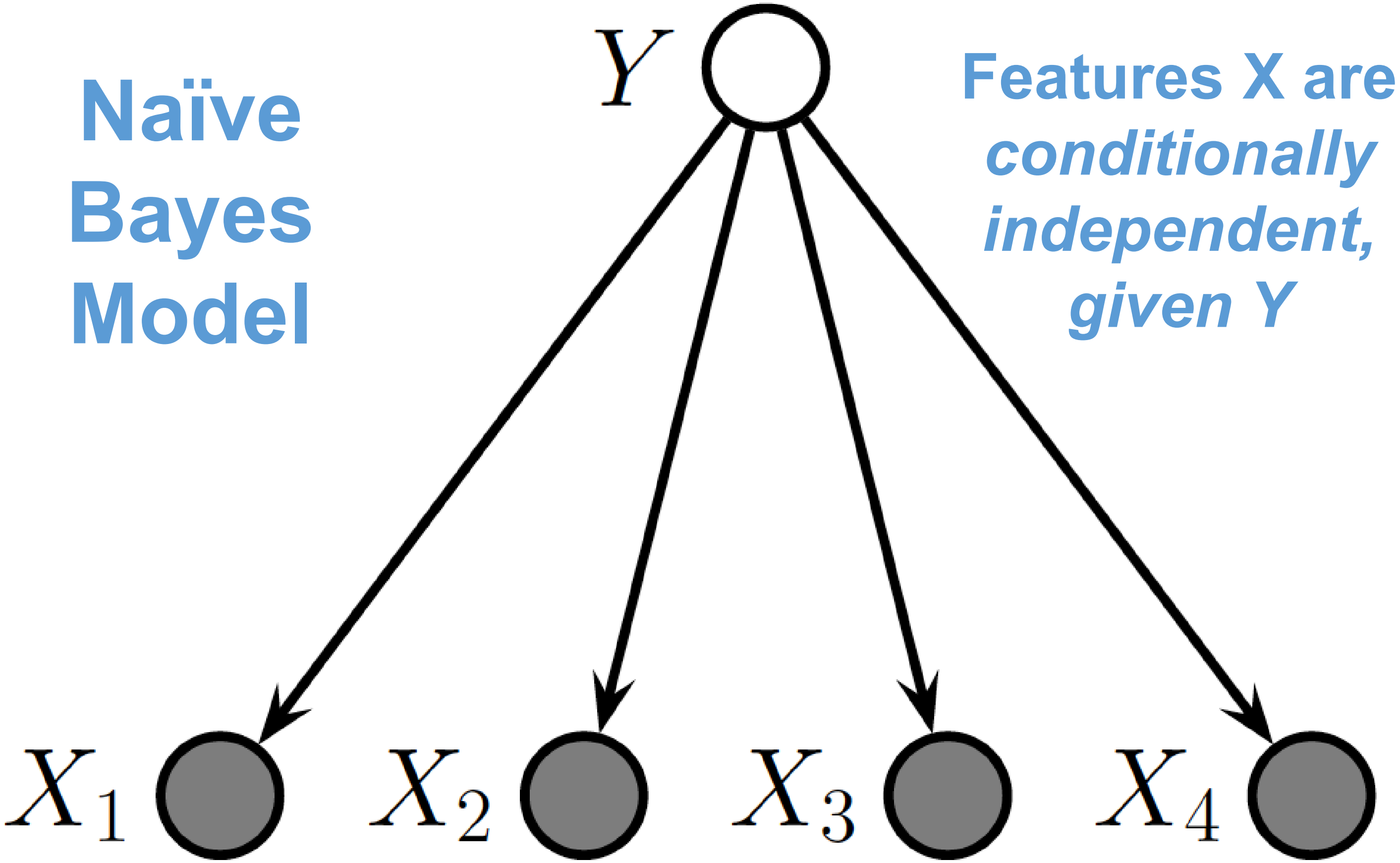
$$p(x) = p(x_3)p(x_1 \mid x_3)p(x_4 \mid x_3)p(x_2 \mid x_1, x_3)$$

Locally normalized factors yield *globally normalized* joint probability

Shading & Plate Notation

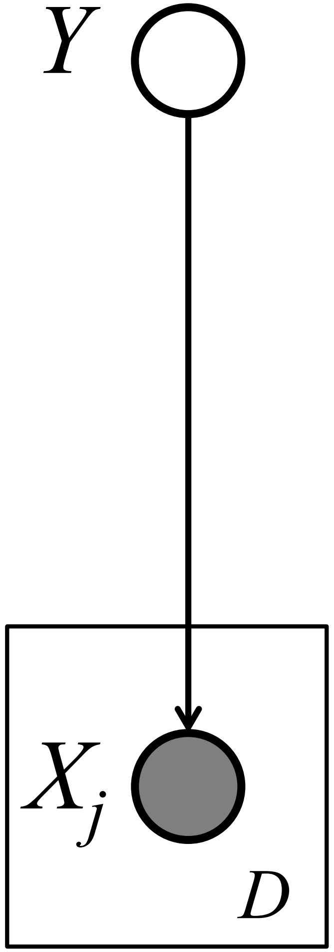
Convention: *Shaded nodes are observed, open nodes are latent/hidden/unobserved*

Naïve Bayes Model



Features X are *conditionally independent, given Y*

$$p(y, \mathbf{x}) = p(y) \prod_{j=1}^{D} p(x_j | y)$$

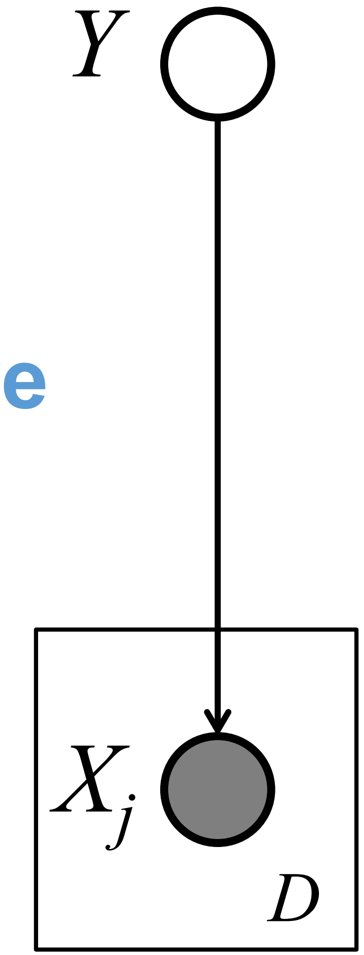


Plates denote replication of random variables

Inference

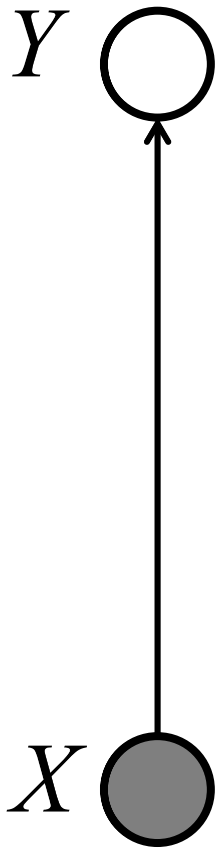
Interpret inference as <u>inverting arrows</u> in the graphical model

Naïve Bayes Generative Model



$$p(y,\mathbf{x}) = p(y)\prod_{j=1}^{D} p(x_j|y)$$

Posterior Model



Posterior

Marginal Likelihood

$$p(y,\mathbf{x}) = p(y \mid \mathbf{x})p(\mathbf{x})$$

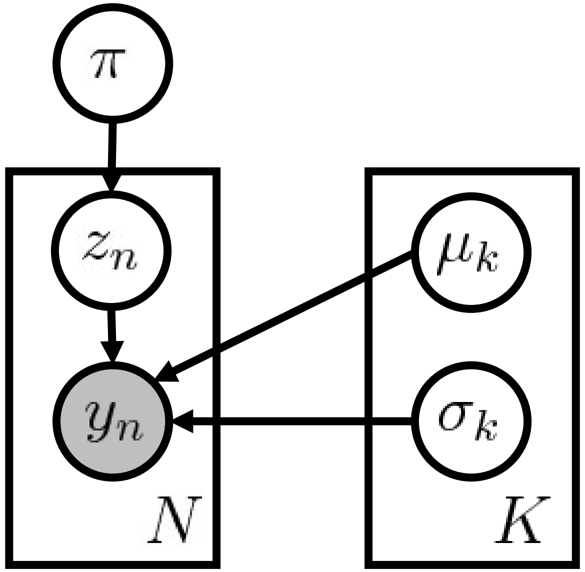
Example: Gaussian Mixture Model

Bayes nets are easily simulated via ancestral sampling…

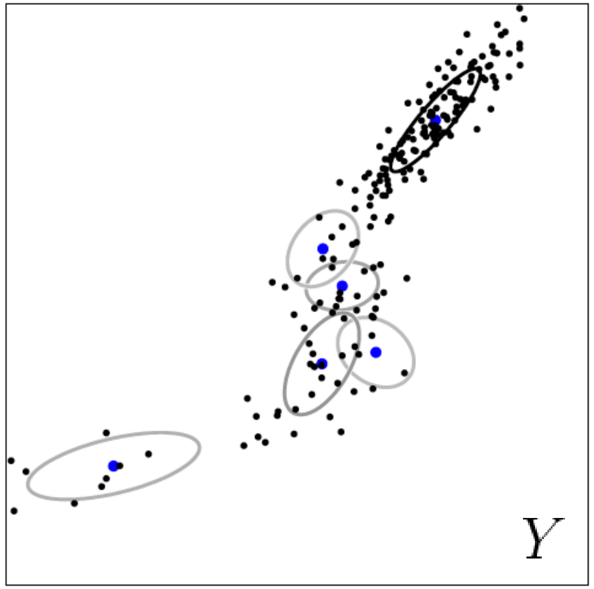
Probability Model

$$
\begin{aligned}
\pi &\sim \mathrm{Dirichlet}(\cdot) \\
\mu_k &\sim \mathcal{N}(\cdot) \\
\sigma_k &\sim \text{Inv-Gamma}(\cdot) \\
z_n \mid \pi &\sim \mathrm{Cat}(\pi) \\
y_n \mid z_n, \mu_{z_n}, \sigma_{z_n} &\sim \mathcal{N}(\mu_{z_n}, \sigma_{z_n})
\end{aligned}
$$

Bayes Net

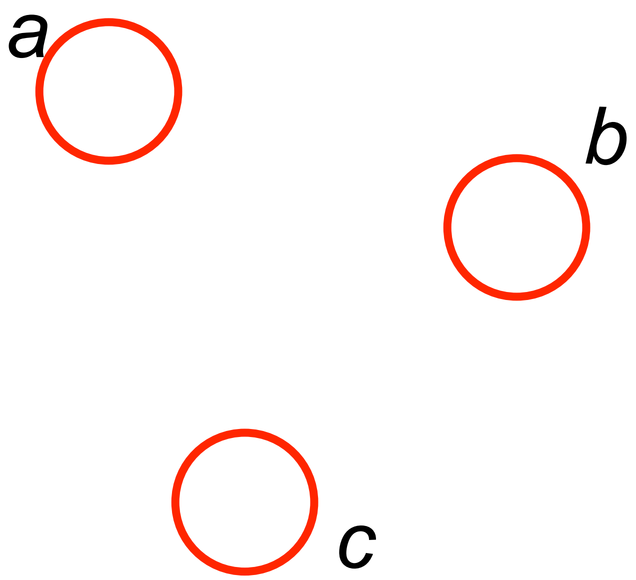


Joint Sample

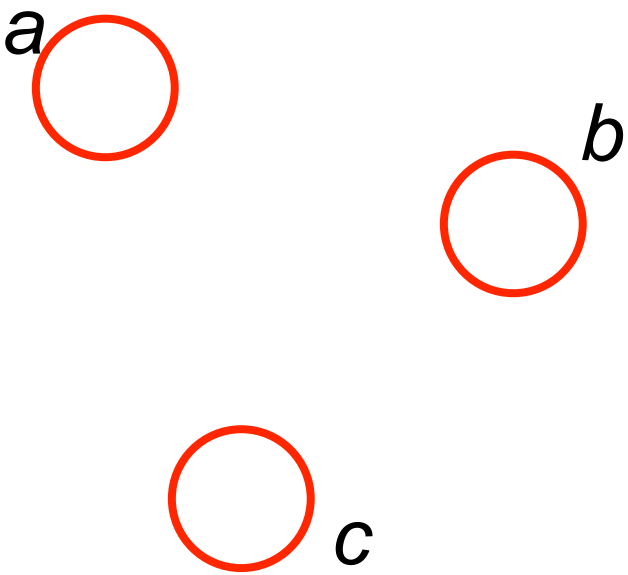


Sample all nodes with no parents, then children, etc., to terminals. Can sample nodes at same level in parallel.

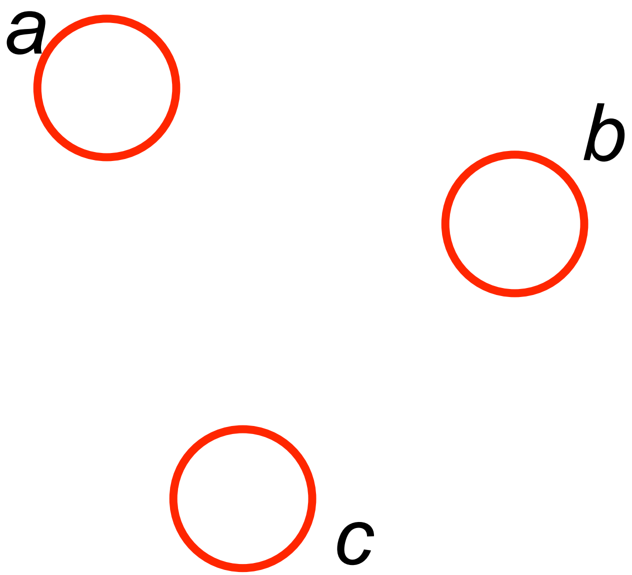
What is the joint factorization?



$$p(a,b,c) = p(a)p(b)p(c)$$

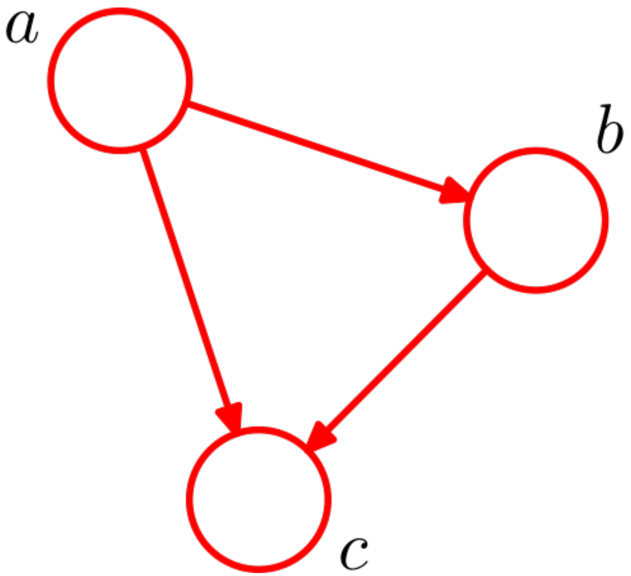


Are a and b independent ($a \perp b$)?



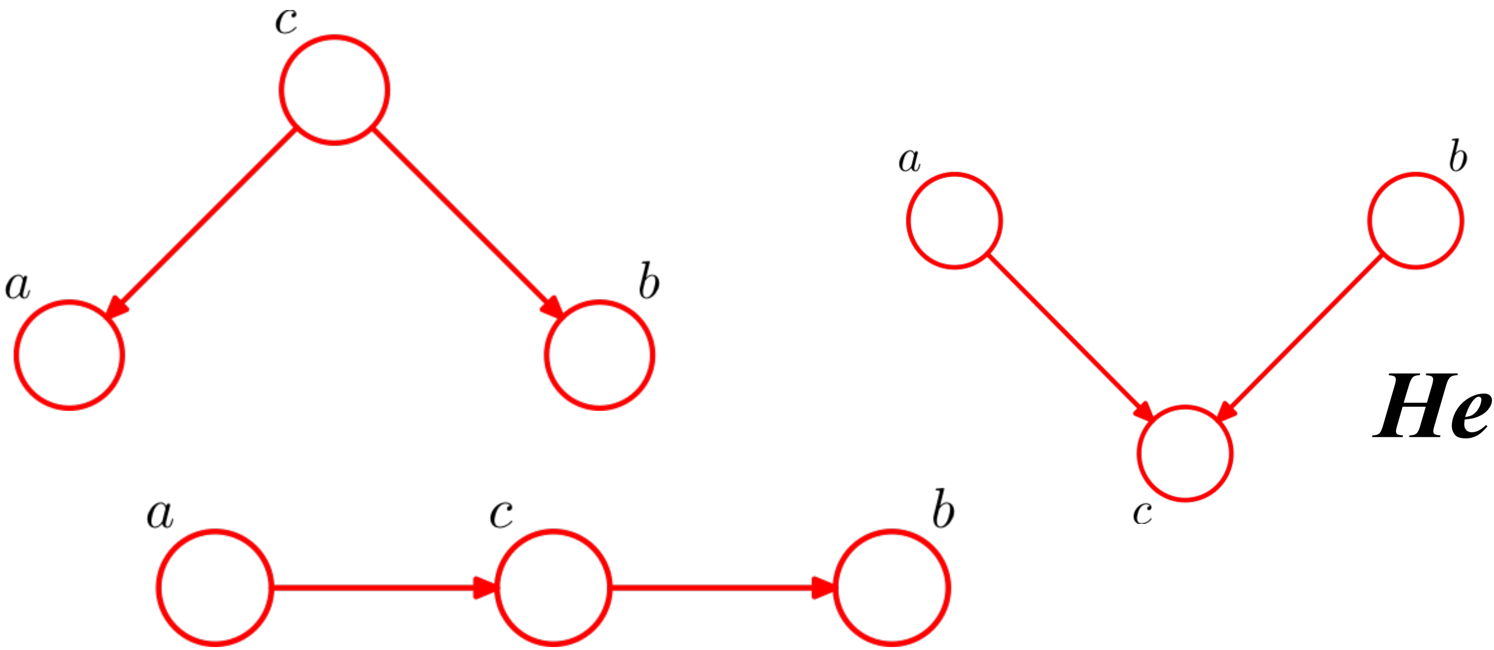
$$\mathbf{p(a,b,c) = p(a)p(b)p(c)}$$

$$p(a,b,c) = p(a)p(b|a)p(c|a,b)$$



Note there are **no conditional independencies** (fully connected graph)

Three interesting cases

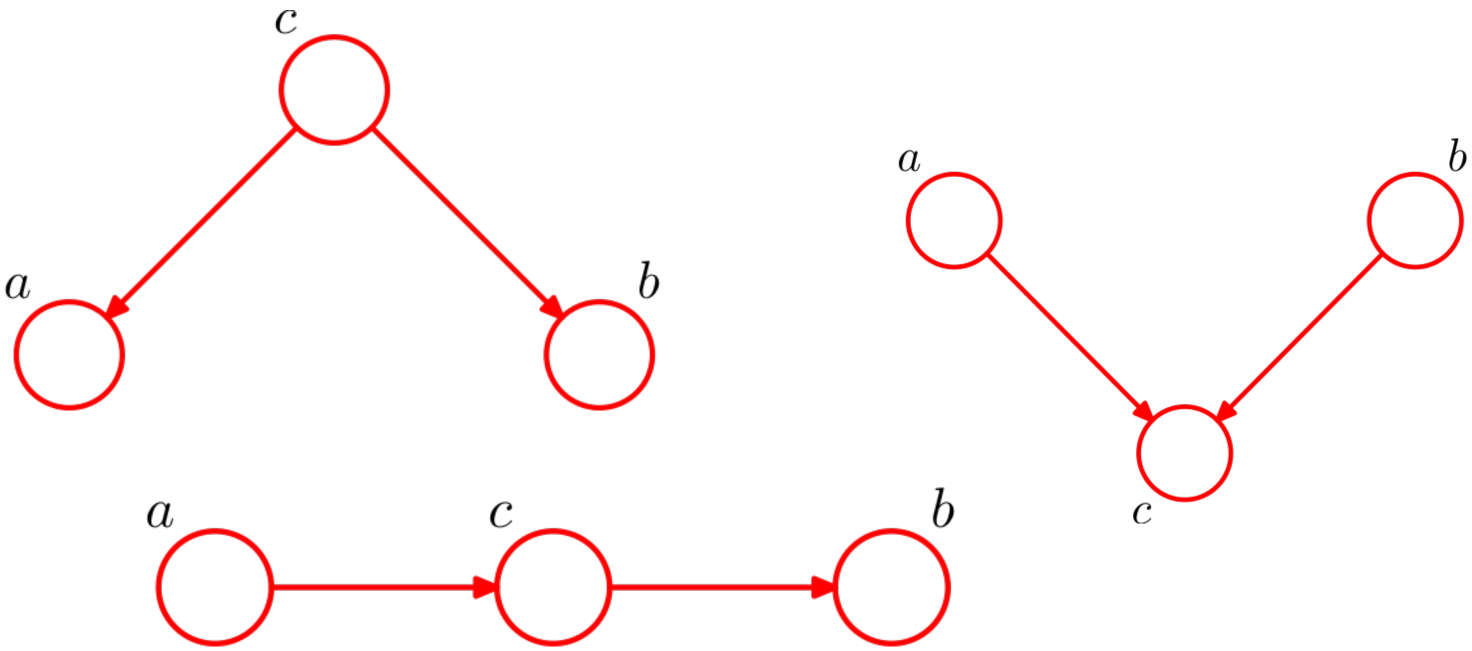


Tail-to-tail

Head-to-head

Head-to-tail

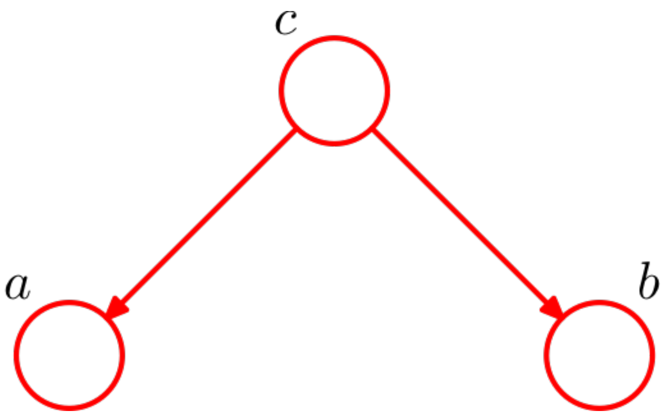
Three interesting cases



For each case, consider two questions:

1) Is $a \perp b$?
2) Is $a \perp b \mid c$? (i.e. c is observed)

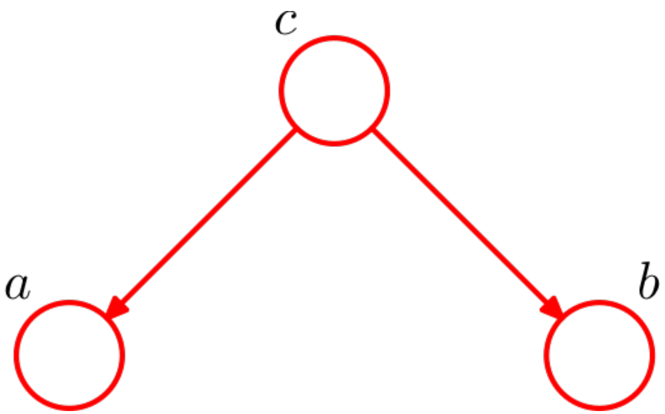
Case one (tail-to-tail)



$$a \not\perp b$$

If you know *a*, that informs you about *c* (by Bayes) which informs you about *b*.

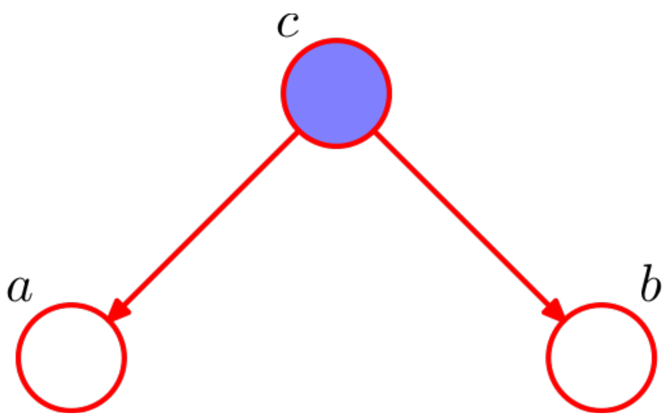
Case one (tail-to-tail)



$$a \not\perp\!\!\!\perp b$$

We can prove this intuitive claim with a counter example.
(HW question)

Case one where c is observed



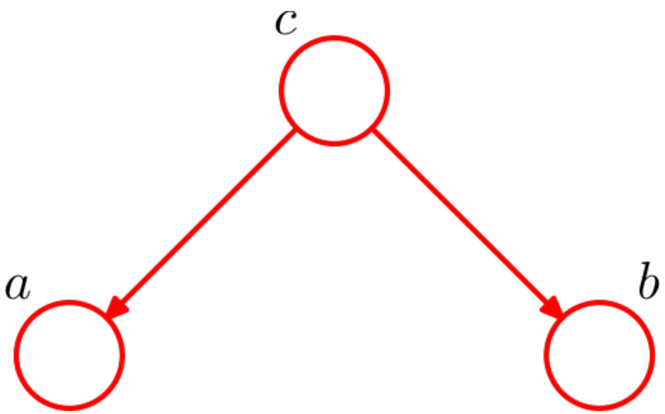
$$a \perp b \mid c$$

$$p(a,b,c) = p(c)p(a|c)p(b|c) \qquad \text{(what the graph represents in general)}$$

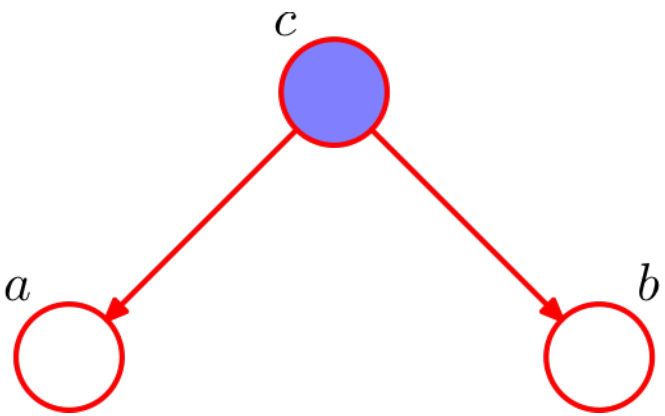
$$p(a,b|c) = p(a|c)p(b|c) \qquad \text{(with } c \text{ observed)}$$

This is the definition of $a \perp b|c$

Case one (tail-to-tail) summary



$$a \not\perp b$$

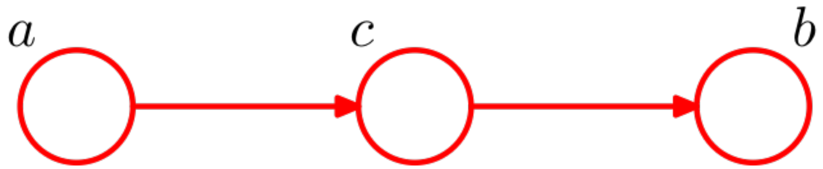


$$a \perp b \mid c$$

Tail-to-tail case

- With no conditioning, no independence
- With conditioning, we have independence

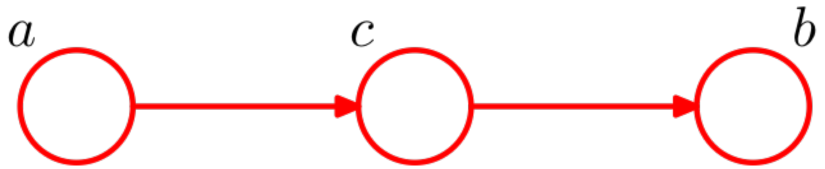
Case two (head-to-tail)



$$a \not\perp b$$

If you know a, that informs you about c, which informs you about b.

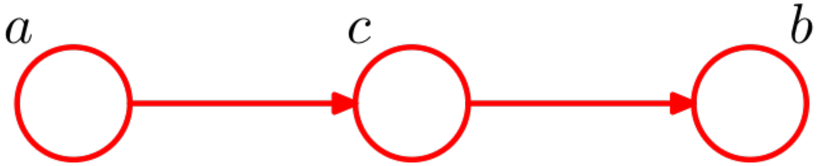
Case two (head-to-tail)



$$a \not\perp b$$

The graph represents $p(a,b,c) = p(a)\,p(c|a)\,p(b|c)$

Case two (head-to-tail)



$$a \not\perp b$$

The graph represents $p\left(a,b,c\right) = p\left(a\right)p\left(c\middle|a\right)p\left(b\middle|c\right)$

Algebraically,

$$p\left(a,b\right) = \sum_{c} p\left(a,b,c\right) = p\left(a\right)\sum_{c} p\left(c\middle|a\right)p\left(b\middle|c\right)$$

If $a \perp b$ then the above would also have to be equal to $p(a)p(b)$

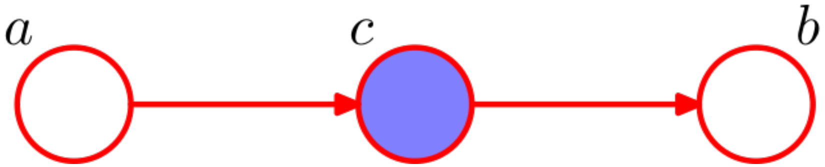
$$p\left(a,b\right) = \sum_{c} p\left(a,b,c\right) = p\left(a\right) \sum_{c} p\left(c \middle| a\right) p\left(b \middle| c\right)$$

If $a \perp b$ then the above **also** equals $p(a)p(b)$

To prove the claim that a $\not\perp$ b we can construct a counter example where the above is false.

Homework Question

Case two where c is observed

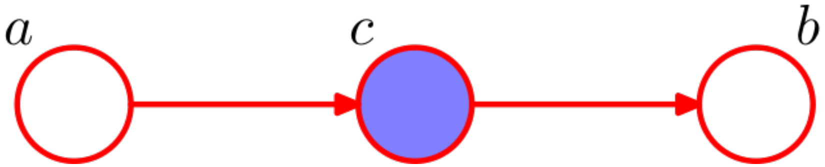


$$a \perp b \mid c$$

$$p\left(a,b \mid c\right) = \frac{p\left(a,b,c\right)}{p\left(c\right)}$$

(why?)

Case two where c is observed

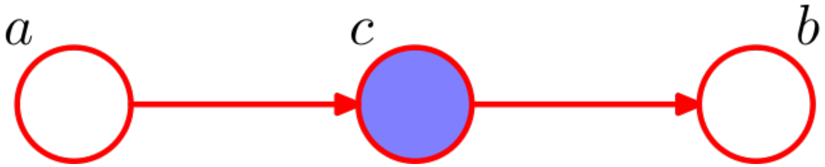


$$a \perp b \mid c$$

$$p(a,b \mid c) = \frac{p(a,b,c)}{p(c)} \qquad \text{(definition)}$$

$$= \frac{p(a)\,p(c|a)\,p(b|c)}{p(c)} \qquad \text{(why?)}$$

Case two where c is observed



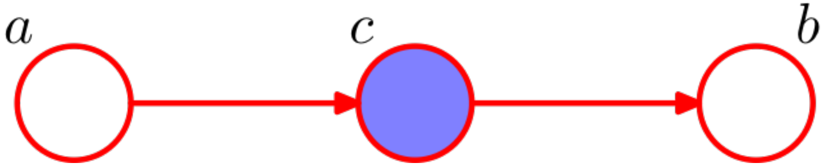
$$a \perp b \mid c$$

$$p(a,b \mid c) = \frac{p(a,b,c)}{p(c)} \qquad \text{(definition)}$$

$$= \frac{p(a)p(c|a)p(b|c)}{p(c)} \qquad \text{(from graph)}$$

$$= \frac{p(a)p(a|c)p(c)p(b|c)}{p(a)p(c)} \qquad \text{(why?)}$$

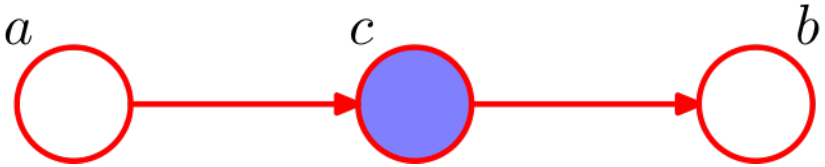
Case two where c is observed



$$a \perp b \mid c$$

$$
\begin{aligned}
p(a,b \mid c) &= \frac{p(a,b,c)}{p(c)} && \text{(definition)} \\
&= \frac{p(a)\,p(c|a)\,p(b|c)}{p(c)} && \text{(from graph)} \\
&= \frac{p(a)\,p(a|c)\,p(c)\,p(b|c)}{p(a)\,p(c)} && \text{(Bayes on } p(c|a)\text{)} \\
&= p(a|c)\,p(b|c) && \text{(why?)}
\end{aligned}
$$

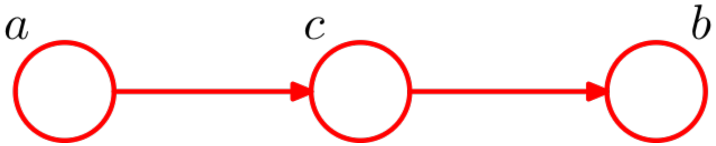
Case two where c is observed



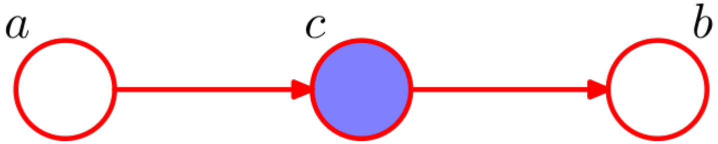
$a \perp b \mid c$ ✓

$$
\begin{aligned}
p\left(a,b \mid c\right) &= \frac{p\left(a,b,c\right)}{p\left(c\right)} && \text{(definition)} \\
&= \frac{p\left(a\right)p\left(c \middle| a\right)p\left(b \middle| c\right)}{p\left(c\right)} && \text{(from graph)} \\
&= \frac{p\left(a\right)p\left(a \middle| c\right)p\left(c\right)p\left(b \middle| c\right)}{p\left(a\right)p\left(c\right)} && \text{(Bayes on } p\left(c \middle| a\right)\text{)} \\
&= p\left(a \middle| c\right)p\left(b \middle| c\right) && \text{(canceling factors)}
\end{aligned}
$$

Case two (head-to-tail) summary



$$a \not\perp b$$



$$a \perp b \mid c$$

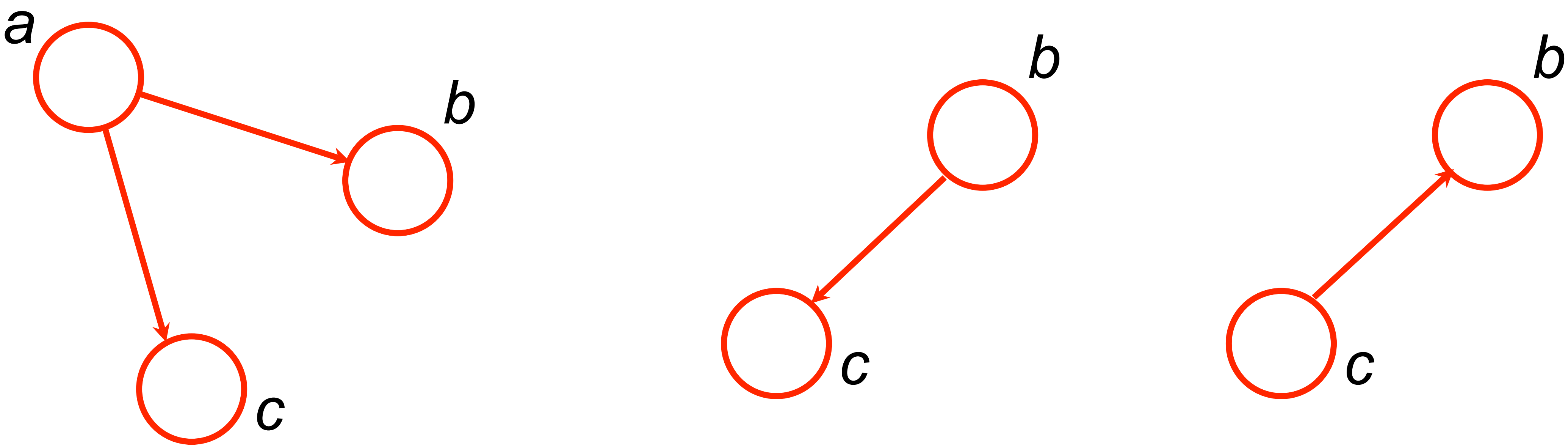
Head-to-tail case

- With no conditioning, no independence
- With conditioning, we have independence

(Same as case one!)

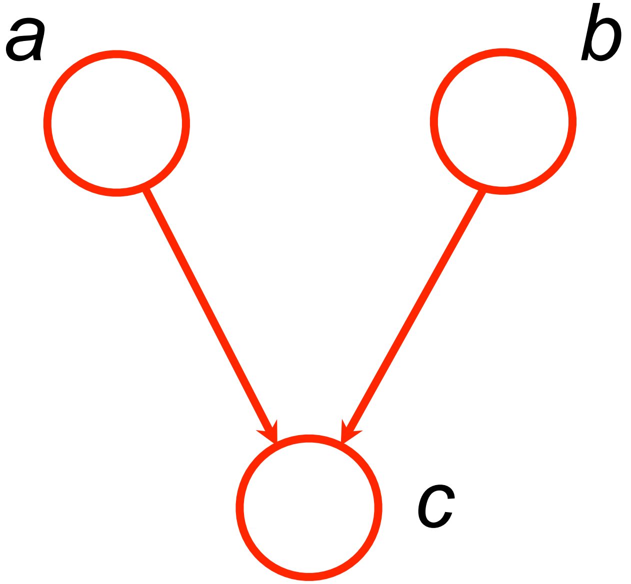
Are b and c independent ($b \perp c$)?

$$p(b,c) = \sum_a p(a)p(b \mid a)p(c \mid a) = p(b)p(c \mid b) = p(c)p(b \mid c)$$



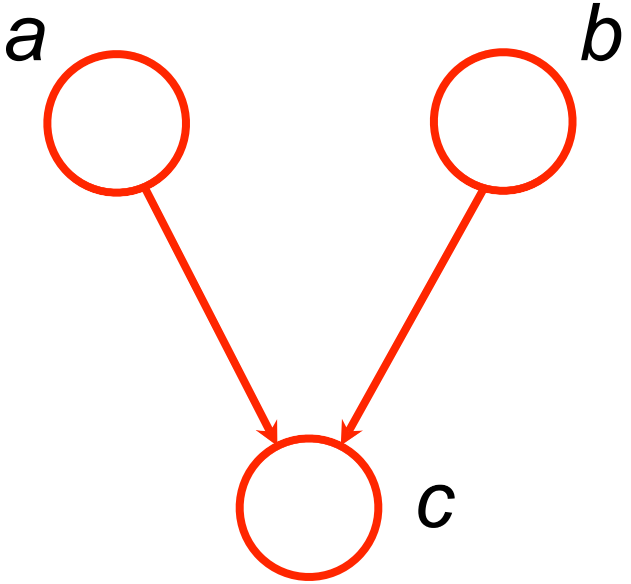
$$p(b,c) \neq p(b)p(c)$$

Are a and b independent ($a \perp b$)?



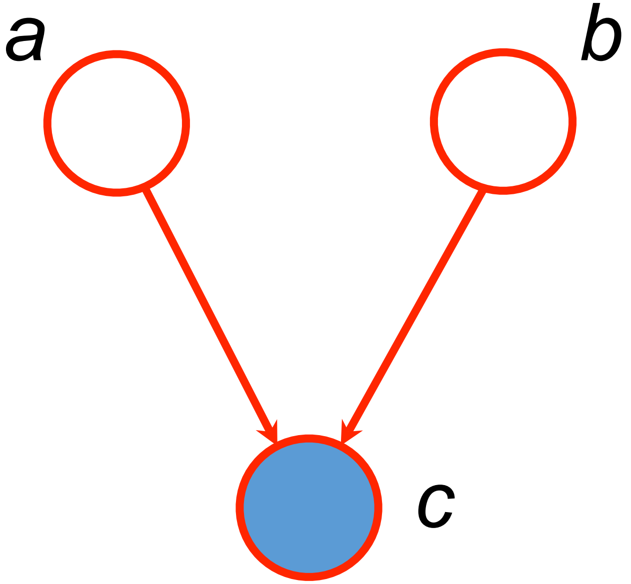
$$p(a,b,c) = p(a)p(b)p(c|a,b)$$

Are a and b independent ($a \perp b$)?



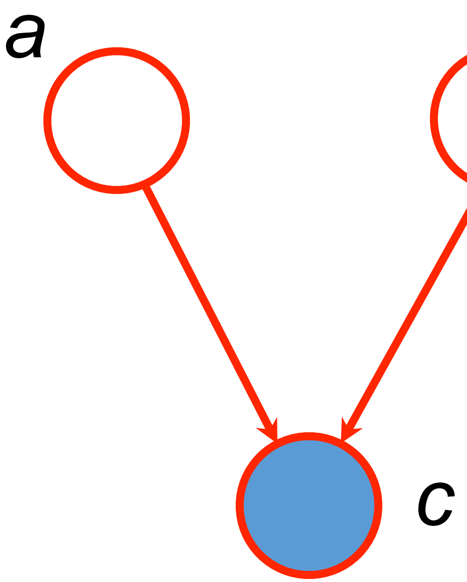
$$p(a,b) = \sum_c p(a)p(b)p(c \mid a,b) = p(a)p(b)$$

Are a and b conditionally independent ($a \perp b \mid c$)?



$$p(a,b,c) = p(a)p(b)p(c|a,b)$$

Are a and b conditionally independent ($a \perp b \mid c$ *)?*

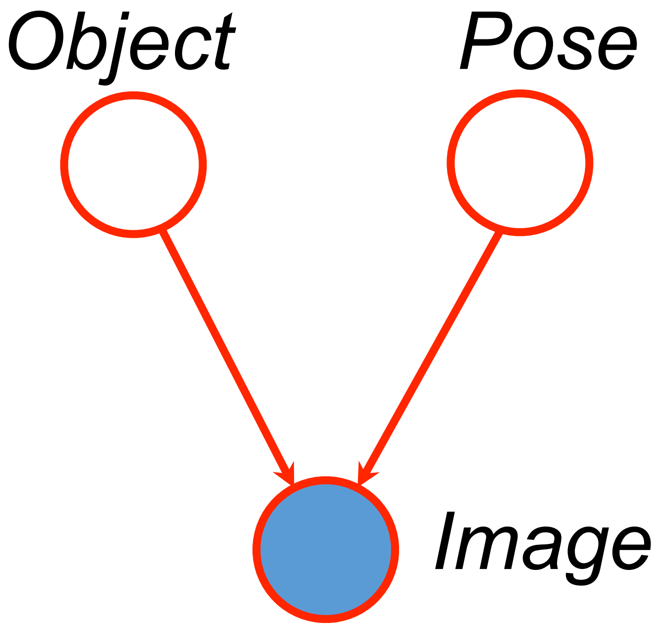


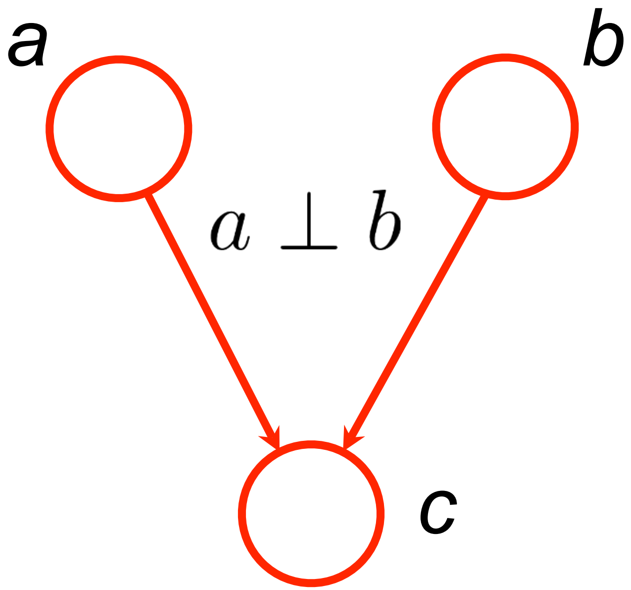
$$
\begin{aligned}
p(a,b|c) &= \frac{p(a,b,c)}{p(c)} \\
&= \frac{p(a)p(b)p(c|a,b)}{p(c)} \\
&\neq p(a|c)p(b|c) \qquad \text{(in general)}
\end{aligned}
$$

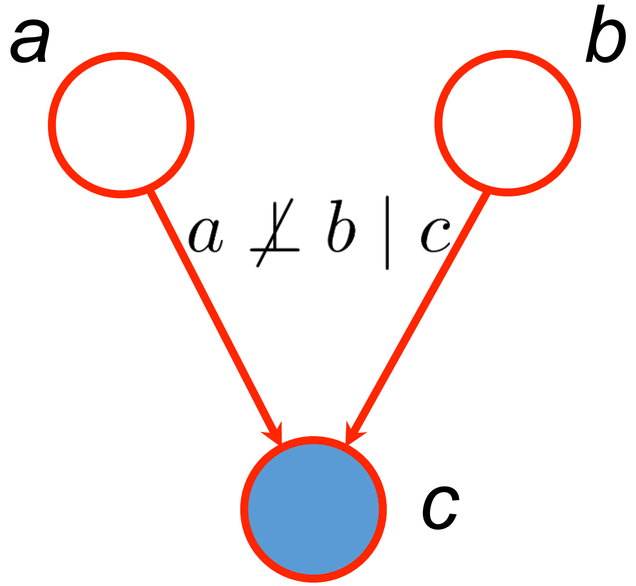
Attempt at algebraic proof.

Unless the algebra reduces to something obviously false, we typically look for a counter example

Both latent variables must explain same observed data so become dependent







Phenomenon in Bayes networks known as **explaining away**

Administrivia

- HW2
  - Will be posted right after class
  - Due Wed, Sep 30, 11:59pm

- HW1: Being graded

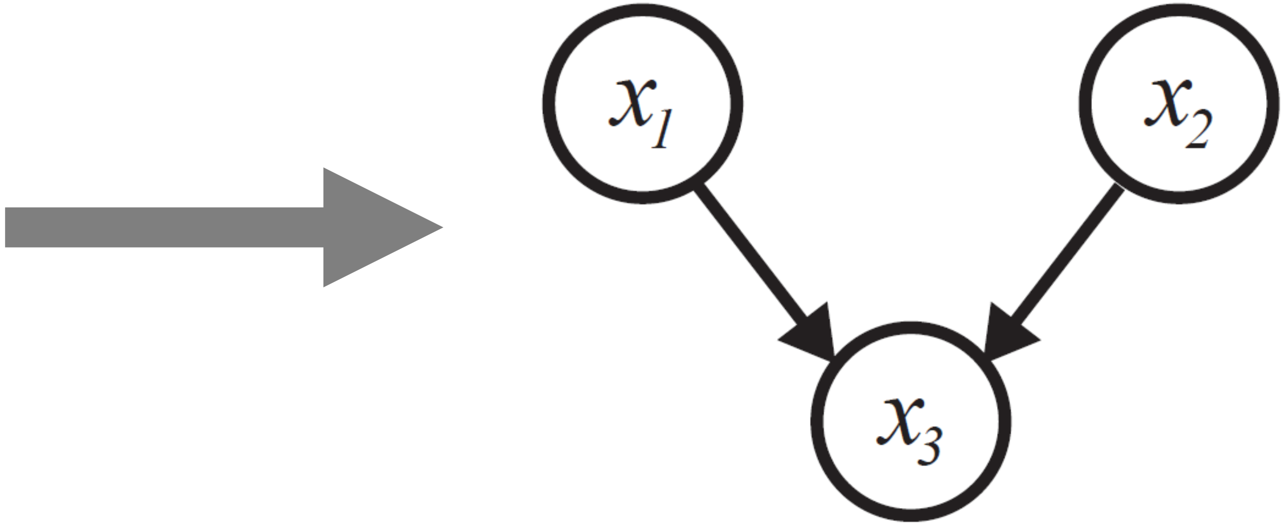
Markov Properties

*How can we be sure a PGM is **correct** for a distribution p(x)?*

Probability Model:

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

Graphical Model:

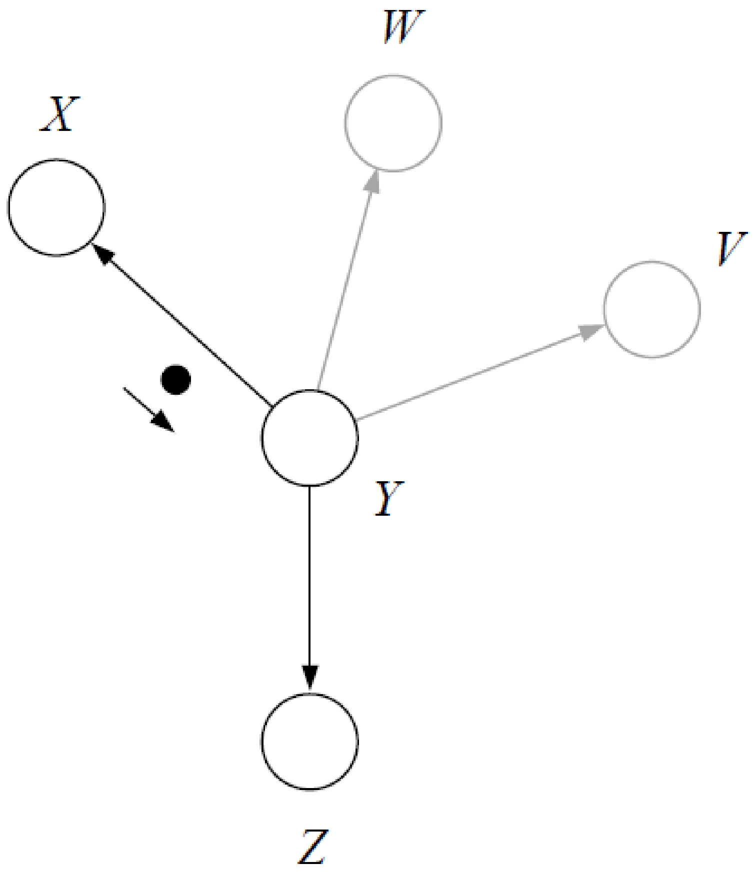


It must satisfy **all** of the conditional independencies of p(x), then we say p(x) **is Markov with respect to** the graph.

Bayes Ball Algorithm

[Source: Michael I Jordan]

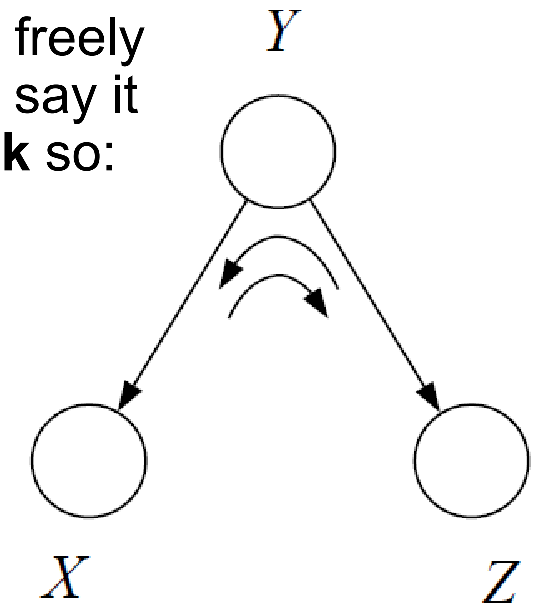
To test if $X \perp Z \mid Y$ imagine rolling a “ball” from X towards Z



The ball follows rules defined by the canonical 3-node subgraphs we’ve discussed

The ball passes freely from X-to-Z, we say it **does not block** so:

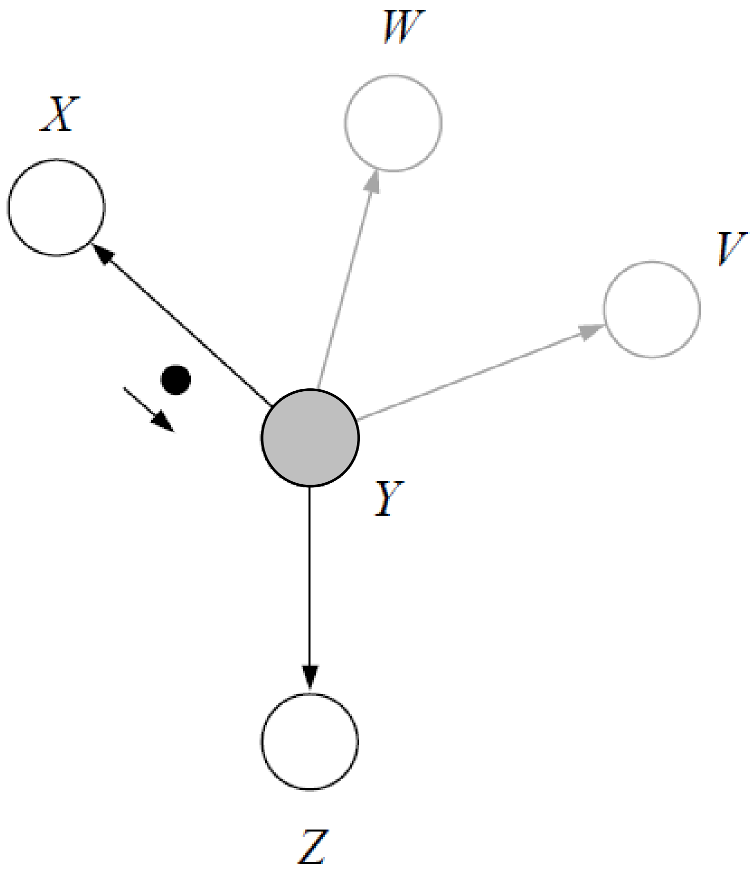
$$X \not\perp Z$$



Bayes Ball Algorithm

[Source: Michael I Jordan]

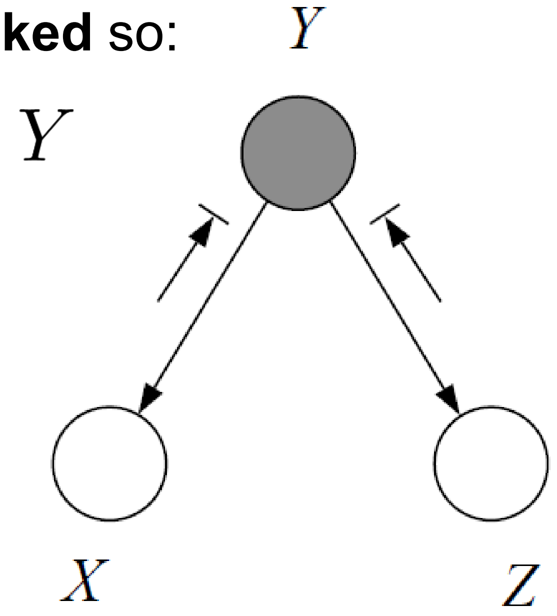
To test if $X \perp Z \mid Y$ imagine rolling a "ball" from X towards Z



The ball follows rules defined by the canonical 3-node subgraphs we've discussed

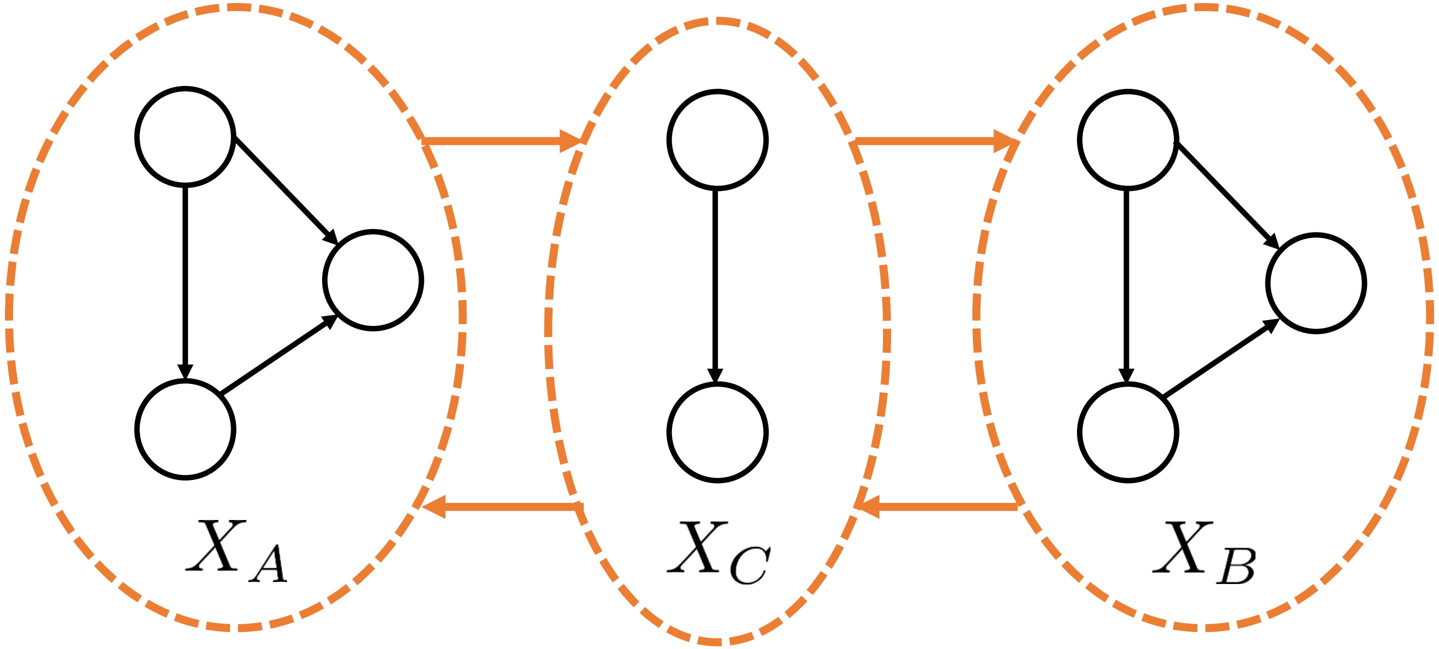
The ball **is blocked** so:

$$X \perp Z \mid Y$$



Bayes Ball Algorithm

To test if $X_A \perp X_B \mid X_C$ roll ball from *every node in* X_A …



If *any* ball reaches *any node* in X_B then

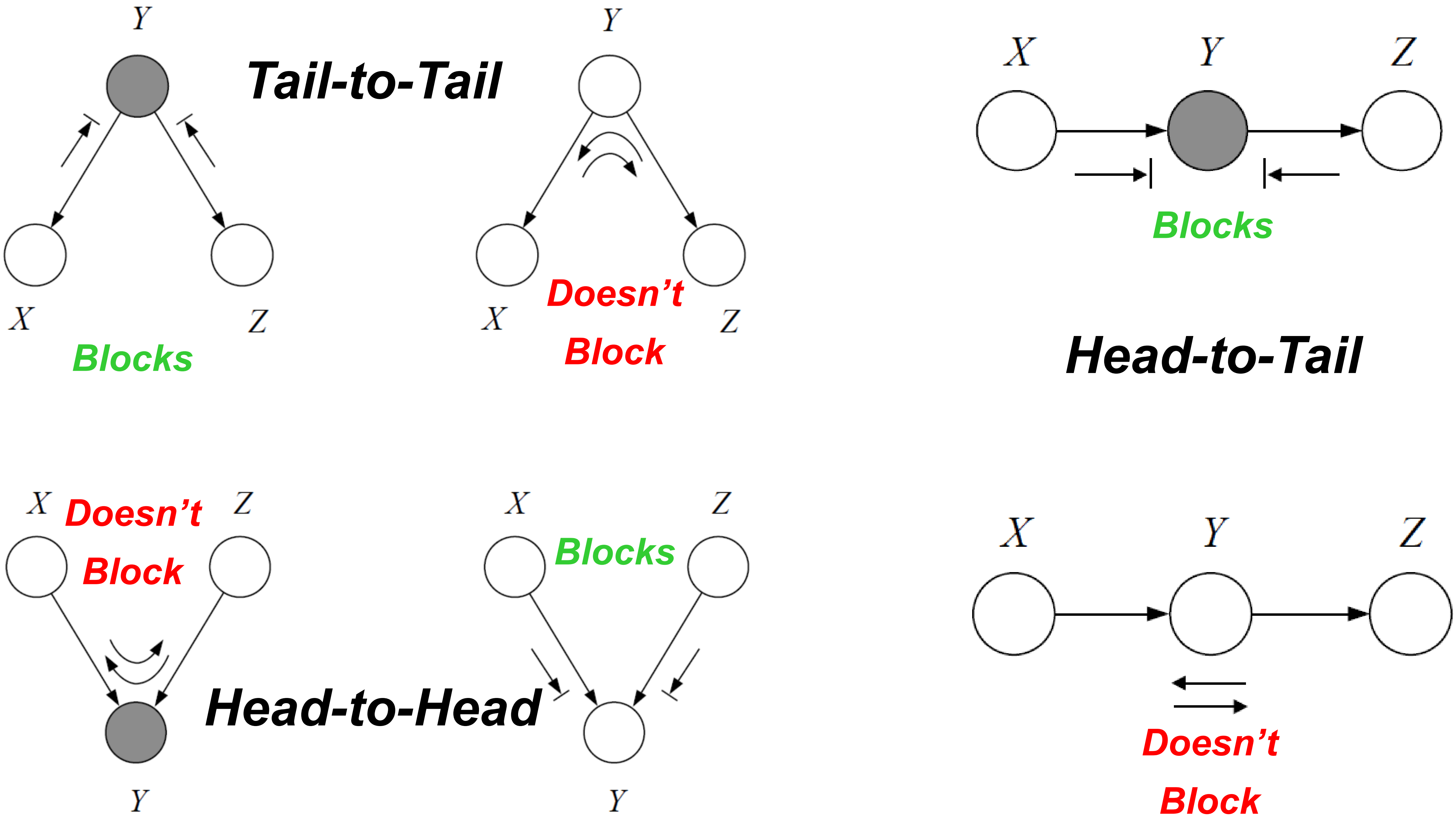
$$X_A \not\perp X_B \mid X_C$$

Otherwise:

$$X_A \perp X_B \mid X_C$$

Tests for property of *directed separation* (d-separation): if X_C separates / blocks X_A from X_B then $X_A \perp X_B \mid X_C$

Bayes Ball Algorithm



Tail-to-Tail

Blocks

Doesn’t Block

Head-to-Tail

Blocks

Doesn’t Block

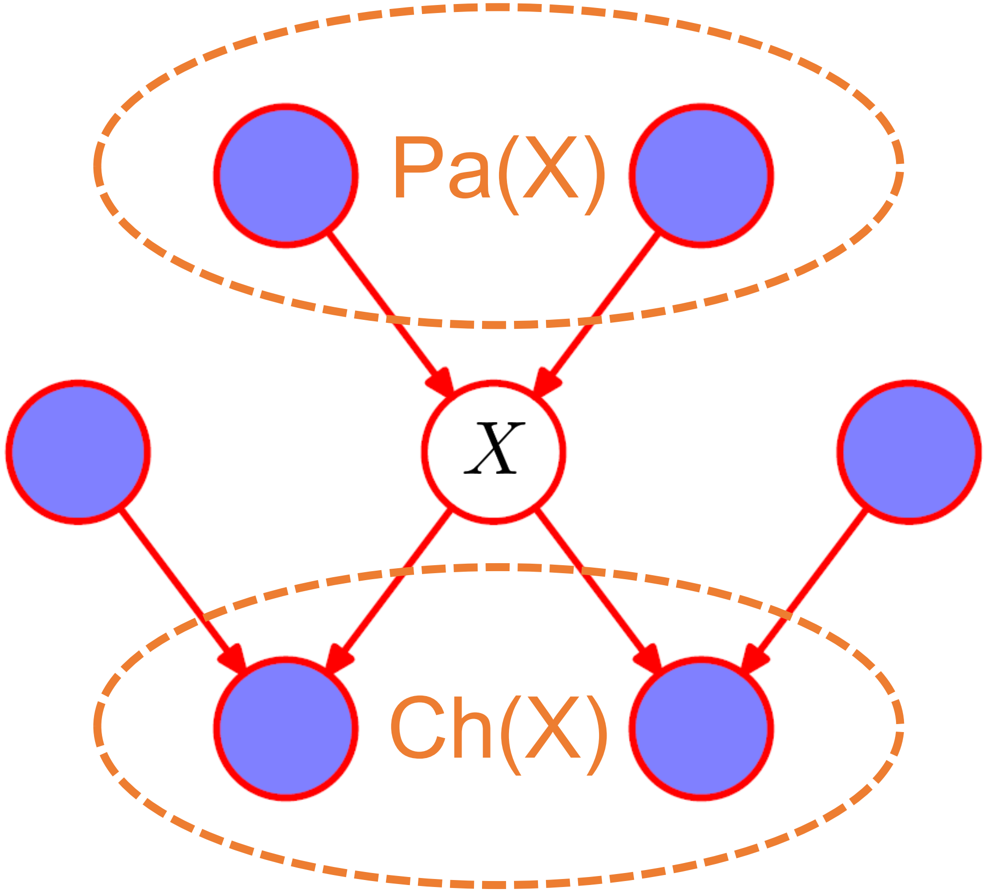
Head-to-Head

Doesn’t Block

Blocks

Markov Blanket

X *conditionally independent of all other nodes, given its Markov blanket*

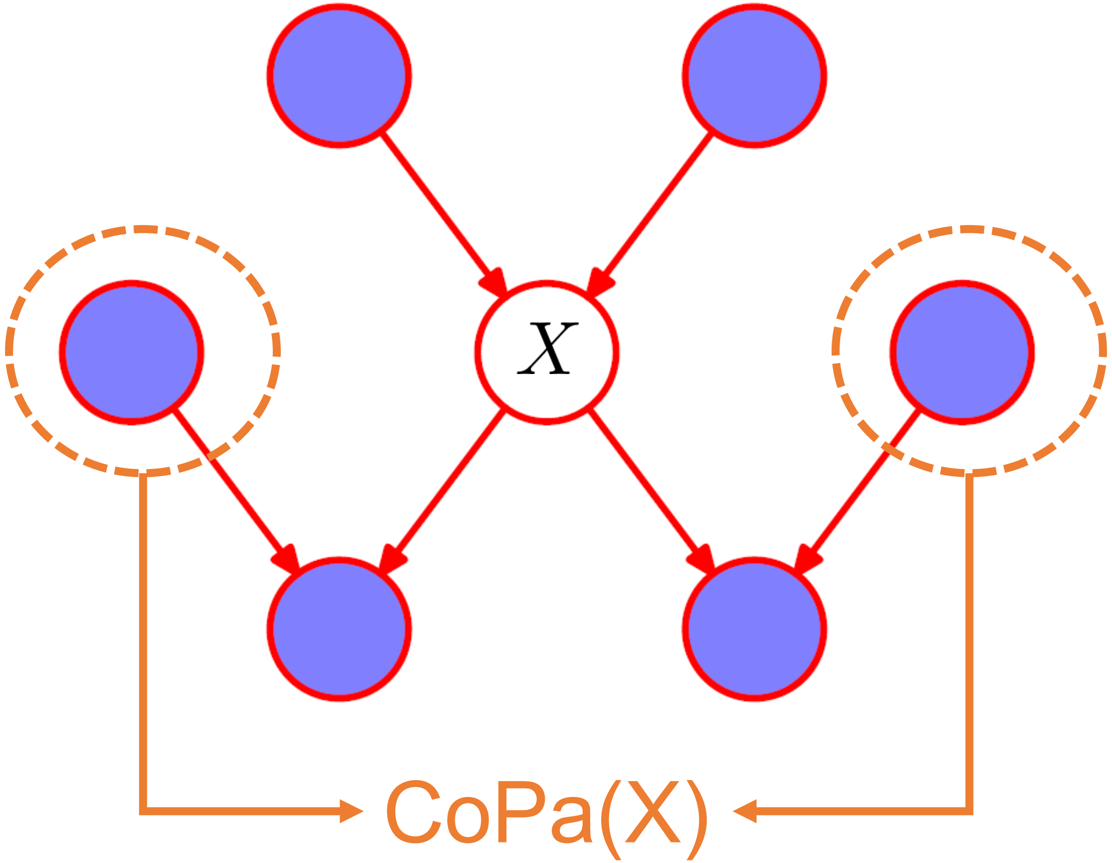


Markov Blanket

X conditionally independent of all other nodes, given its Markov blanket

Q: Why co-parents?

A: Explaining away



Markov Blanket

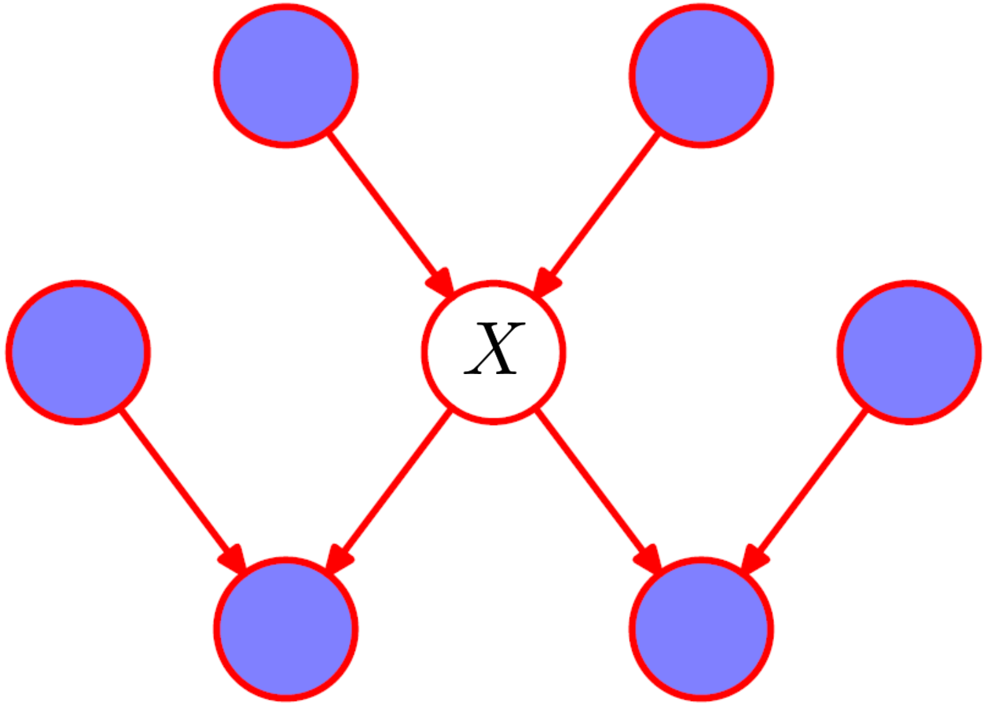
X conditionally independent of all other nodes, given its Markov blanket

Definition *A RV X with distribution p(x) that is Markov w.r.t. graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has a* **Markov blanket** *given by:*

$$\mathrm{Mb}(X) = \mathrm{Pa}(X) \cup \mathrm{Ch}(X) \cup \mathrm{CoPa}(X)$$

For any $Y \notin \mathrm{Mb}(X)$:

$$X \perp Y \mid \mathrm{Mb}(X)$$



Markov blanket used to simplify inference and distribute computation (e.g. Gibbs sampler, variational inference, etc.)

Directed Models Summary

- Distribution factorized as product of conditionals via chain rule

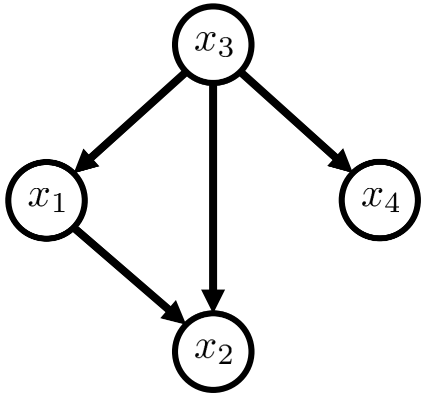
$$p(x_1, x_2, x_3, x_4) = p(x_3)p(x_1 \mid x_3)p(x_4 \mid x_1, x_3)p(x_2 \mid x_1, x_3, x_4)$$

- Choose ordering where terms simplify due to conditional independence

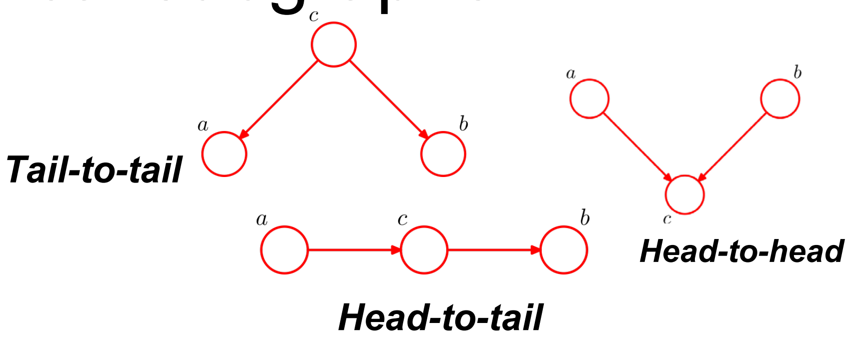
Eg. Suppose $x_4 \perp x_1 \mid x_3$ and $x_2 \perp x_4 \mid x_1$ then:

$$p(x) = p(x_3)p(x_1 \mid x_3)p(x_4 \mid x_3)p(x_2 \mid x_1, x_3)$$

- Directed graph encodes factorized distribution via conditional independence properties



- Test independence using canonical subgraphs:
- Straightforward simulation via **ancestral sampling**



Outline

Directed graphical models

- Bayes Nets
- Conditional dependence

Undirected graphical models

- Markov random fields (MRFs)
- Factor graphs

Factorized Probability Distributions

A probability distribution over RVs $x = (x_1, \dots, x_d)$ can be written as a product of factors,

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

*A **minimal factorization** is one where all factors are **maximal cliques** (not a strict subset of any other clique) in the MRF*

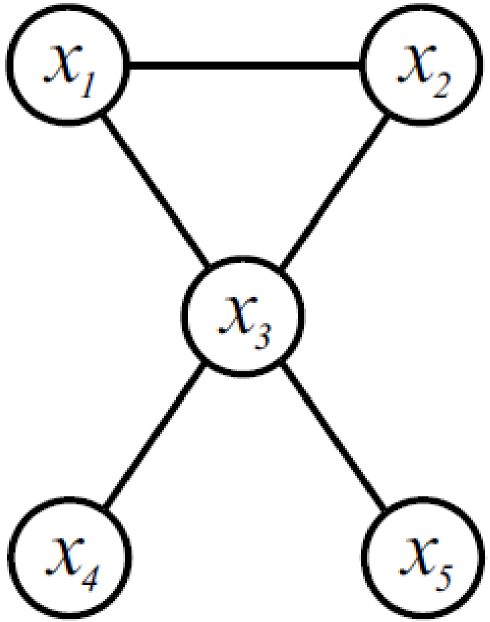
Where:

- $\mathcal{C}$ a collection of subsets of indices $\{1, \dots, d\}$
- $\psi(\cdot)$ are nonnegative *factors* (or *potential functions*)
- Z the normalizing constant (or *partition function*)

$$Z = \int \prod_{c \in \mathcal{C}} \psi_c(x_c) \, dx_c$$

Undirected Graphical Models

A **graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a set of vertices $\mathcal{V}$ and edges $\mathcal{E}$. An edge $(s,t) \in \mathcal{E}$ connects two vertices $s,t \in \mathcal{V}$.



In **undirected models** edges are specified irrespective of node ordering so that,

$$(s,t) \in \mathcal{E} \Leftrightarrow (t,s) \in \mathcal{E}$$

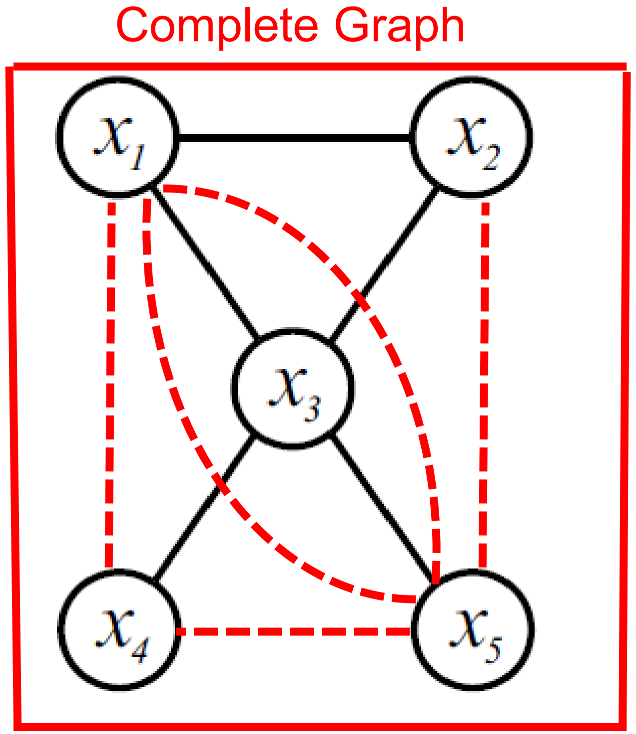
Distributions are typically specified with unknown normalization (easier to specify),

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

Markov Random Fields (MRFs)

A factor $\psi_c(x_c)$ corresponds to a clique $c \in \mathcal{C}$ (fully connected subgraph) in the MRF

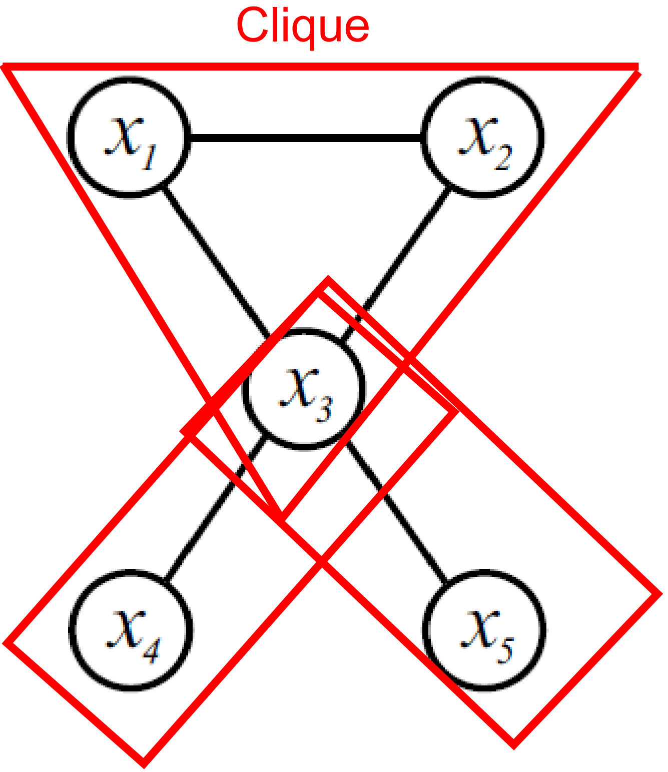
An MRF does not imply a unique factorization, for example all the following are "*valid*":



$$\psi(x_1, x_2, x_3, x_4, x_5)$$

Markov Random Fields (MRFs)

A factor $\psi_c(x_c)$ corresponds to a clique $c \in \mathcal{C}$ (fully connected subgraph) in the MRF



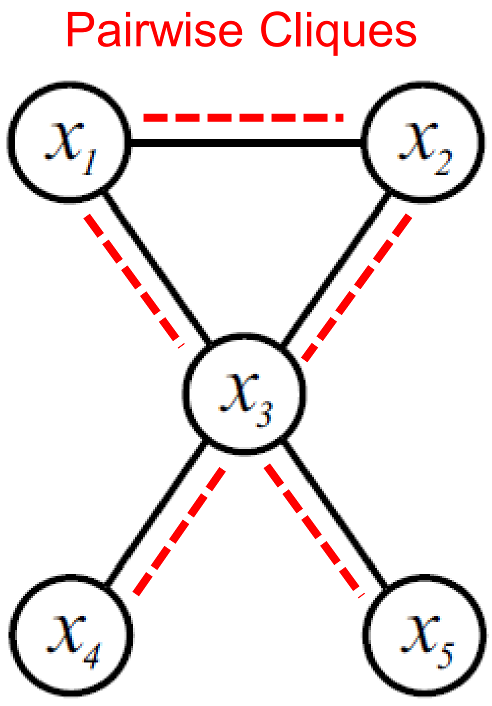
An MRF does not imply a unique factorization, for example all the following are "*valid*":

$$\psi(x_1, x_2, x_3, x_4, x_5)$$

$$\psi(x_1, x_2, x_3)\psi(x_3, x_4)\psi(x_3, x_5)$$

Markov Random Fields (MRFs)

A factor $\psi_c(x_c)$ corresponds to a clique $c \in \mathcal{C}$ (fully connected subgraph) in the MRF



An MRF does not imply a unique factorization, for example all the following are "*valid*":

$$\psi(x_1, x_2, x_3, x_4, x_5)$$

$$\psi(x_1, x_2, x_3)\psi(x_3, x_4)\psi(x_3, x_5)$$

$$\psi(x_1, x_2)\psi(x_2, x_3)\psi(x_1, x_3)\psi(x_3, x_4)\psi(x_3, x_5)$$

A factorization is *valid* if it satisfies the *Global Markov property*, defined by conditional independencies

Conditional Independence (Undirected)

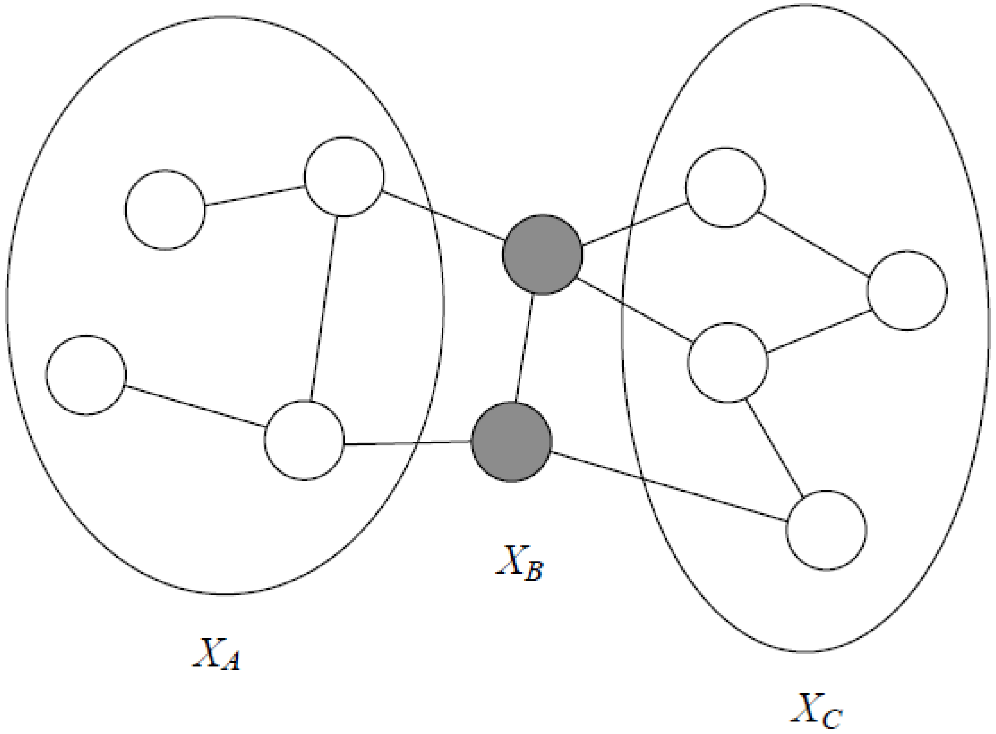
We say x_A and x_C are *independent* or $x_A \perp\!\!\!\perp x_C$ if:

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

We say they are *conditionally independent* or $x_A \perp\!\!\!\perp x_C \mid x_B$ if:

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

Def. We say $p(x)$ is *globally Markov* w.r.t. $\mathcal{G}$ if $x_A \perp\!\!\!\perp x_C \mid x_B$ in any separating set of $\mathcal{G}$.



Conditional independence in undirected graphical models is defined by separating sets

[Source: Michael I. Jordan]

Global & Local Markov Properties

Global Markov Property

- Set B separates A from C if all paths from A to C pass through B
- By definition, distribution is Markov if and only if for any B separating A and C:

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

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

Local Markov Property

- Given its *neighbors*, each node is independent of all other variables

$$p(x_s \mid x_{\mathcal{V} \setminus s}) = p(x_s \mid x_{\Gamma(s)})$$

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

Markov blanket only includes immediate neighbors (we needed co-parents in Bayes nets)

- This local Markov property is a special case of the global Markov property

[Source: Erik Sudderth]

Hammersley-Clifford Theorem

Thorem (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 non-negative potential functions on those cliques is then always Markov with respect to $\mathcal{G}$:*

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

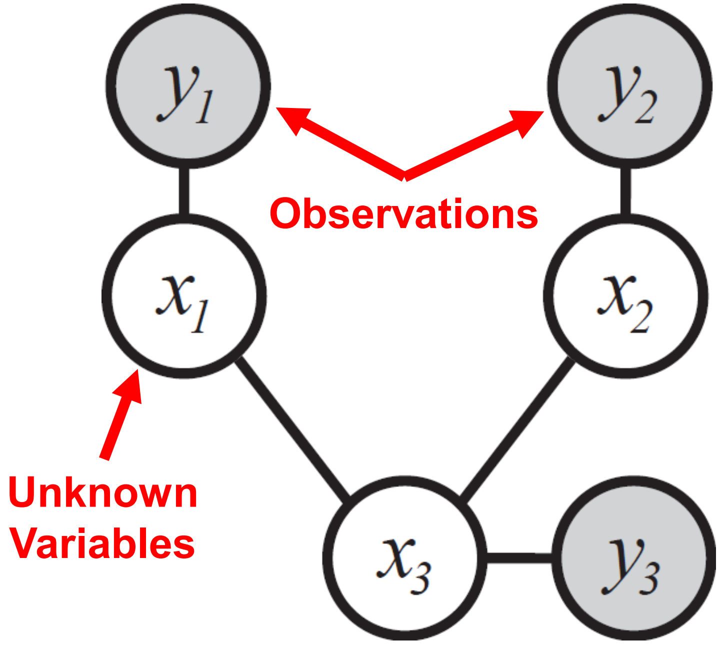
Conversely, any strictly positive density which is Markov with respect to $\mathcal{G}$ can be represented in this factored form.

Global Markov Property ⟷ **Joint Factorization**

(Graph Separation Implies Independence) (Potential Function for Each Clique)

Pairwise Markov Random Field

Often easier to specify and do inference on pairwise model

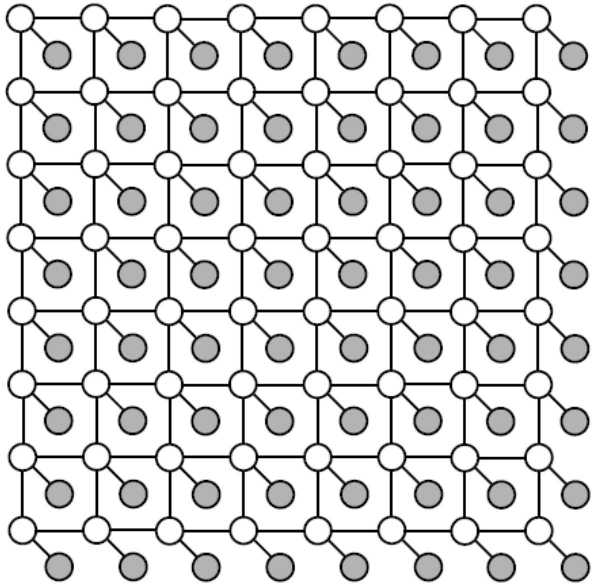


$$p(x,y) \propto \prod_{s\in\mathcal{V}} \underbrace{\psi_s(x_s,y)}_{\textbf{Likelihood}} \prod_{(s,t)\in\mathcal{E}} \underbrace{\psi_{st}(x_s,x_t)}_{\textbf{Prior}}$$

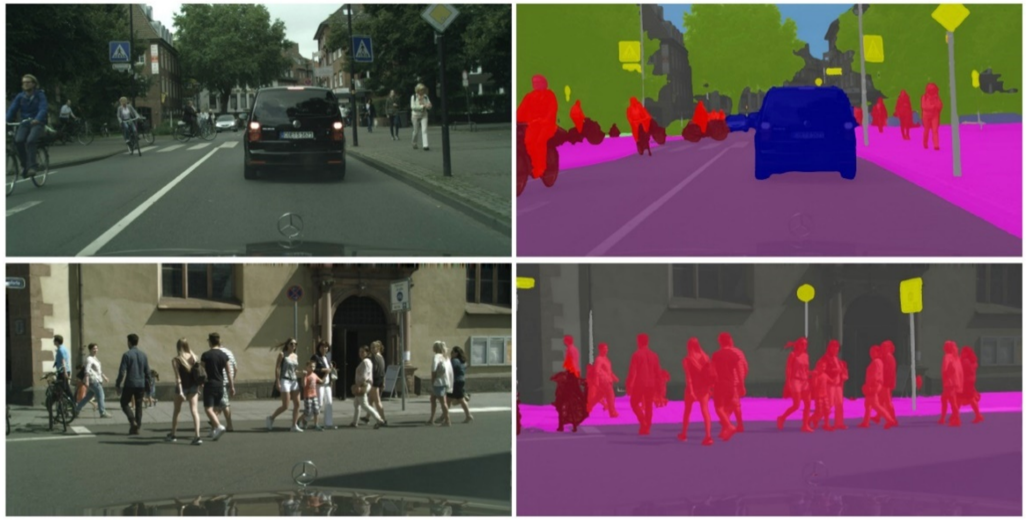
Restricted class of MRFs

- 2-node factor exists for every edge
- Explicit factorization of joint distribution
- High-order factors not always easily decomposed into pairwise terms

Example: Image Segmentation



[Source: Kundu, A. et al., CVPR16]



Don't need to know log-partition to specify model

Pairwise MRF *energy*: $-\log p(x,y) = \log Z + \sum_i \psi_i(x_i, y_i) + \sum_{(i,j)} \psi_{i,j}(x_i, x_j)$

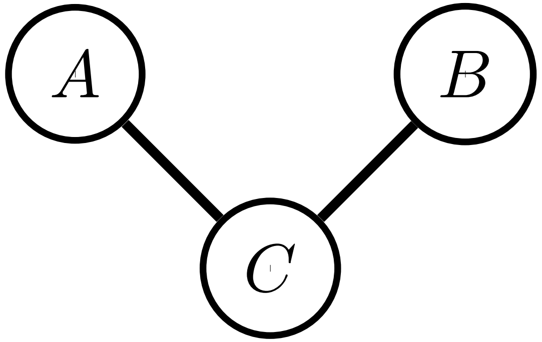
Don't need to specify normalized conditionals as in Bayes Nets

Low energy configurations = High probability

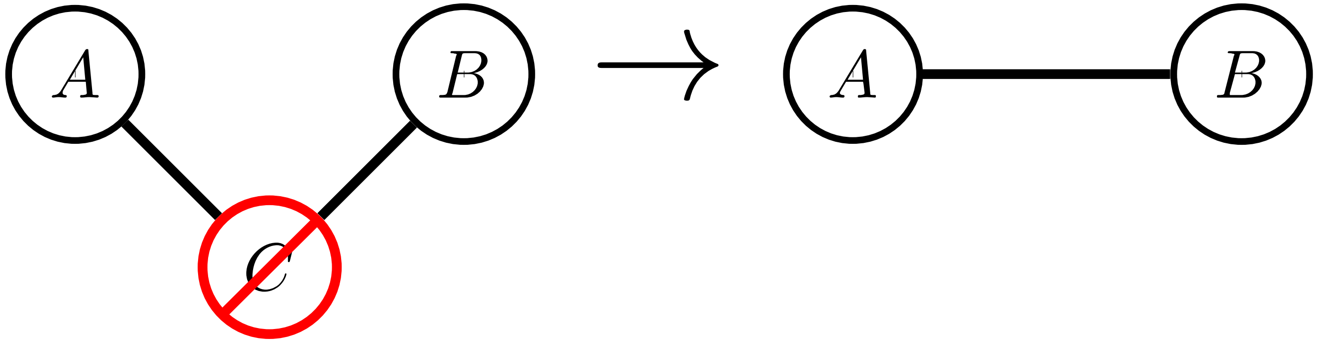
L2 Likelihood: $\psi_i(x_i, y_i) = \|x_i - y_i\|^2$ **Potts model:** $\psi_{i,j}(x_i, x_j) = \mathbb{I}(x_i = x_j)$

MAP (minimum energy) configuration = Piecewise constant regions

Transformations of Undirected Models

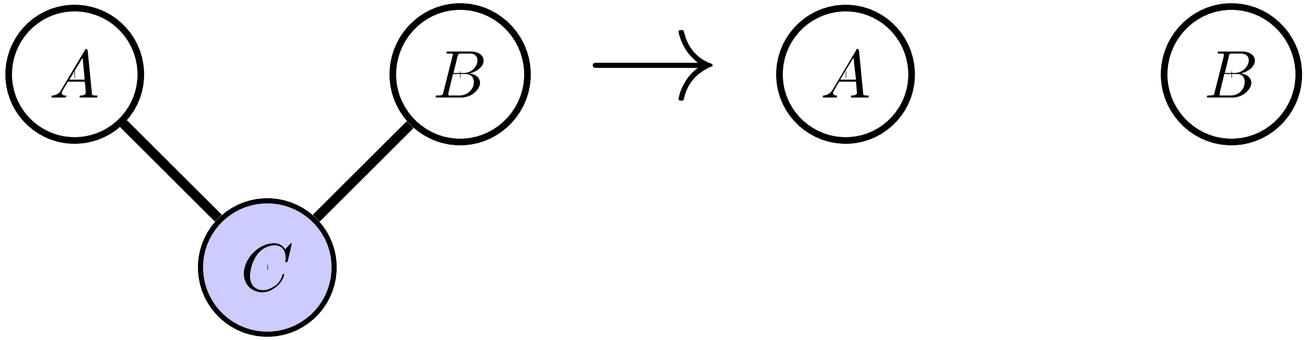


$$p(A, B, C) = \psi_{AC}(A, C)\psi_{BC}(B, Z)/Z$$



$$p(A, B) \neq p(A)p(B)$$

Marginalising over C makes A and B (graphically) dependent.



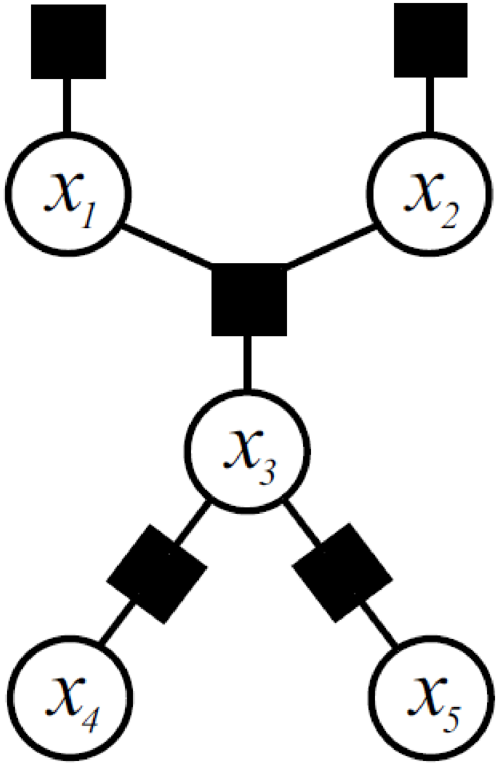
$$p(A, B|C) = p(A|C)p(B|C)$$

Conditioning on C makes A and B independent:

[Source: Erik Sudderth]

Factor Graphs

A *hypergraph* $\mathcal{H} = (\mathcal{V}, \mathcal{F})$ where a *hyperedge* $f \in \mathcal{F}$ is a subset of vertices $f \subset \mathcal{V}$.



Factor node for each product term in the joint factorization:

Graphical model makes factorization explicit

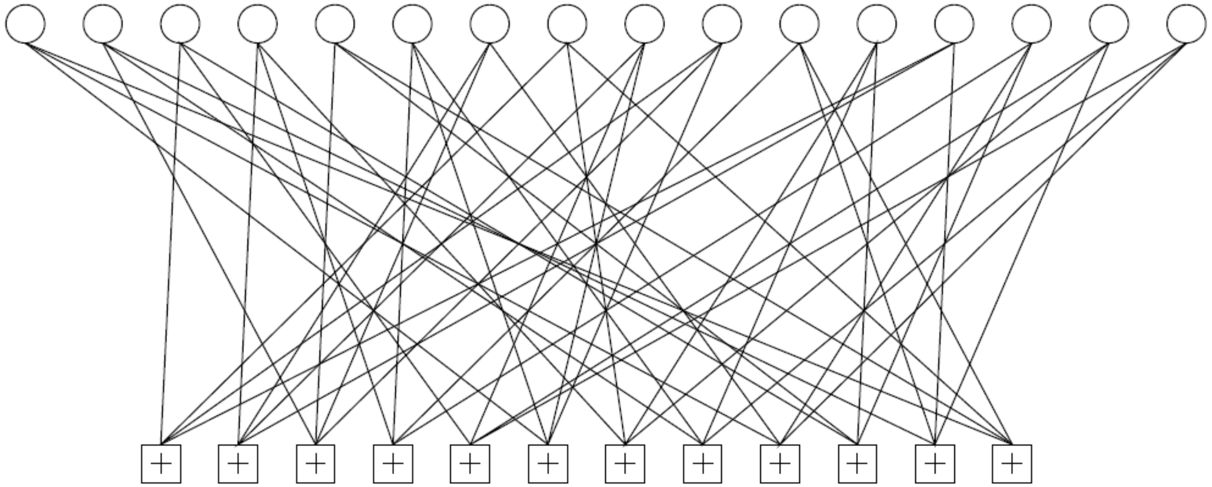
$$p(x) \propto \prod_{f \in \mathcal{F}} \psi_f(x_f)$$

where $x_f = \{x_i : i \in f\}$ the set of variables in factor *f*. For example:

$$\psi(x_1)\psi(x_2)\psi(x_1, x_2, x_3)\psi(x_3, x_4)\psi(x_3, x_5)$$

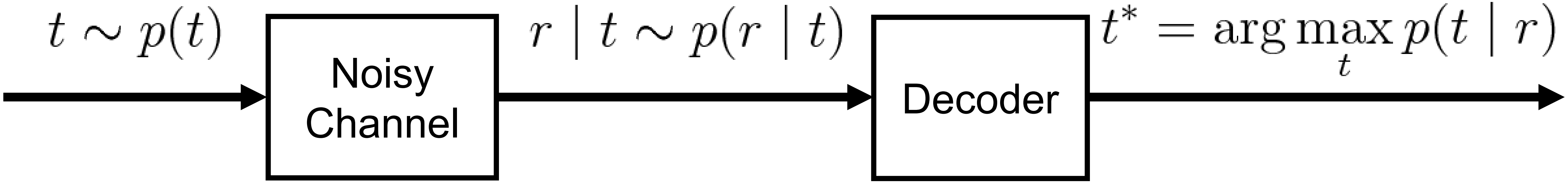
Example: Low Density Parity Check Codes

Factor Graph Representation



Sparse Parity Check Matrix

$$\mathbf{H} = \begin{array}{|cccccccccccccccc|}\hline
 & & 1 & & & & & 1 & & 1 & & & 1 & & & \\
 & & & 1 & & & 1 & & 1 & & & & 1 & & & \\
 & 1 & & & 1 & & 1 & & & 1 & & & & & & \\
 & & & 1 & & 1 & & & & & 1 & & & 1 & & \\
 & & 1 & & & & & & 1 & & & & & & 1 & 1 \\
1 & & & & & 1 & & & 1 & 1 & & & & & & \\
 & & & 1 & & & & 1 & & & & 1 & & & 1 & \\
 & 1 & & & & 1 & & & & & & 1 & & & & 1 \\
1 & & & & & & 1 & & & & & & & 1 & & 1 \\
 & & 1 & & 1 & & & & & & & 1 & & 1 & & \\
 & 1 & & & & & & & & & 1 & & 1 & & 1 & \\
1 & & & & 1 & & & 1 & & & 1 & & & & & \\ \hline
\end{array}$$



Transmitted Code

$t \sim p(t)$

Noisy Channel

Received Code

$r \mid t \sim p(r \mid t)$

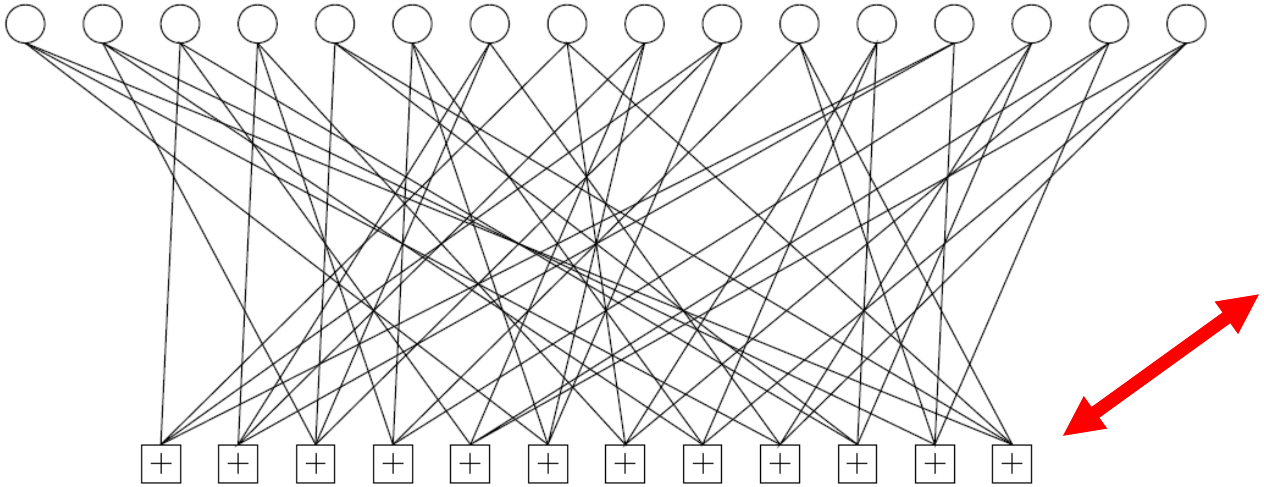
Decoder

$t^* = \arg\max_t p(t \mid r)$

[Source: David MacKay]

Example: Low Density Parity Check Codes

Factor Graph Representation



Sparse Parity Check Matrix

$$\mathbf{H} = \begin{bmatrix}
 & & 1 & & & & & 1 & & 1 & & & 1 & & & \\
 & & & 1 & & & 1 & & 1 & & & & 1 & & & \\
 & 1 & & & 1 & & 1 & & & 1 & & & & & & \\
 & & & 1 & & 1 & & & & & 1 & & & 1 & & \\
 & & 1 & & & & & & 1 & & & & & & 1 & 1 \\
1 & & & & & 1 & & & 1 & 1 & & & & & & \\
 & & & 1 & & & & 1 & & & & 1 & & & 1 & \\
 & 1 & & & & 1 & & & & & & 1 & & & & 1 \\
1 & & & & & & 1 & & & & & & & 1 & & 1 \\
 & & 1 & & 1 & & & & & & & 1 & & 1 & & \\
 & 1 & & & & & & & & & 1 & & 1 & & 1 & \\
1 & & & & 1 & & & 1 & & & 1 & & & & &
\end{bmatrix}$$

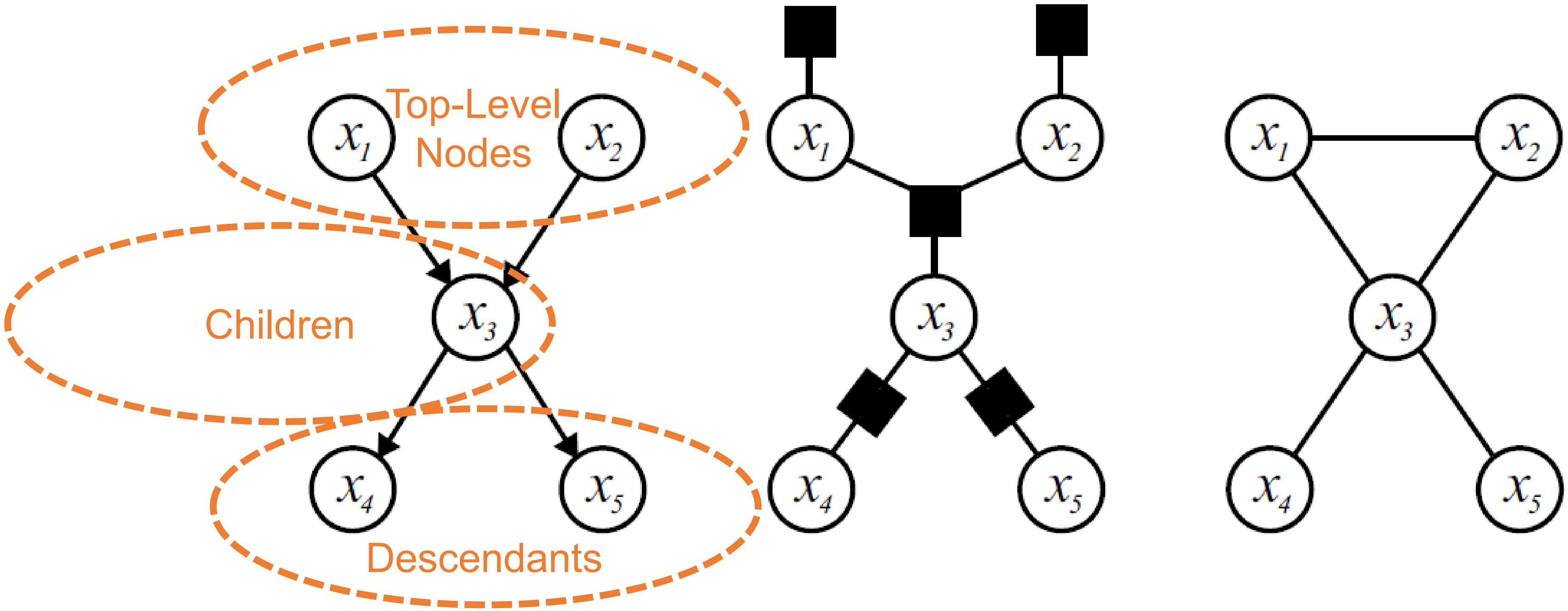
- Valid codes have zero parity: $p(t) \propto \mathbb{I}(Ht = 0 \bmod 2)$
- Chanel noise model arbitrary, e.g. flip bits w/ ϵ probability:

$$p(r \mid t) = \prod_n p(r_n \mid t_n) = \prod_n (1-\epsilon)^{\mathbb{I}(r_n = t_n)} \epsilon^{\mathbb{I}(r_n \neq t_n)}$$

n-th bit →

[Source: David MacKay]

Simulation



Bayes Nets Ancestral sampling successively samples from conditionals:

$$p(\mathbf{x}) = \prod_{i \in \mathcal{V}} p(x_i \mid x_{\mathrm{Pa}(i)}) \qquad \text{so} \qquad x_i \sim p(x_i \mid x_{\mathrm{Pa}(i)})$$

Undirected Graphs Lack locally normalized conditionals to sample from

Undirected Models Summary

- Joint factorization as nonnegative factors (potentials) over subsets:

$$p(x) \propto \prod_{f \in \mathcal{F}} \psi_f(x_f)$$

- Easier to specify models compared to Bayes nets since:
  - Factors do not need to be normalized conditional probabilities
  - May specify up to unknown normalization constant
- Easier to verify Markov independence via *separating sets*
- Factorization ambiguous in MRFs, but explicit in factor graphs (factor graphs generally preferred)
- Simulation is not easy in general. Can't do ancestral sampling.