RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Markov Networks

Edgar Granados

Overview

Markov Networks vs Bayesian Networks

- Graphical Models
- Bayesian Networks - Diagnostics
- Markov Networks - Computer vision

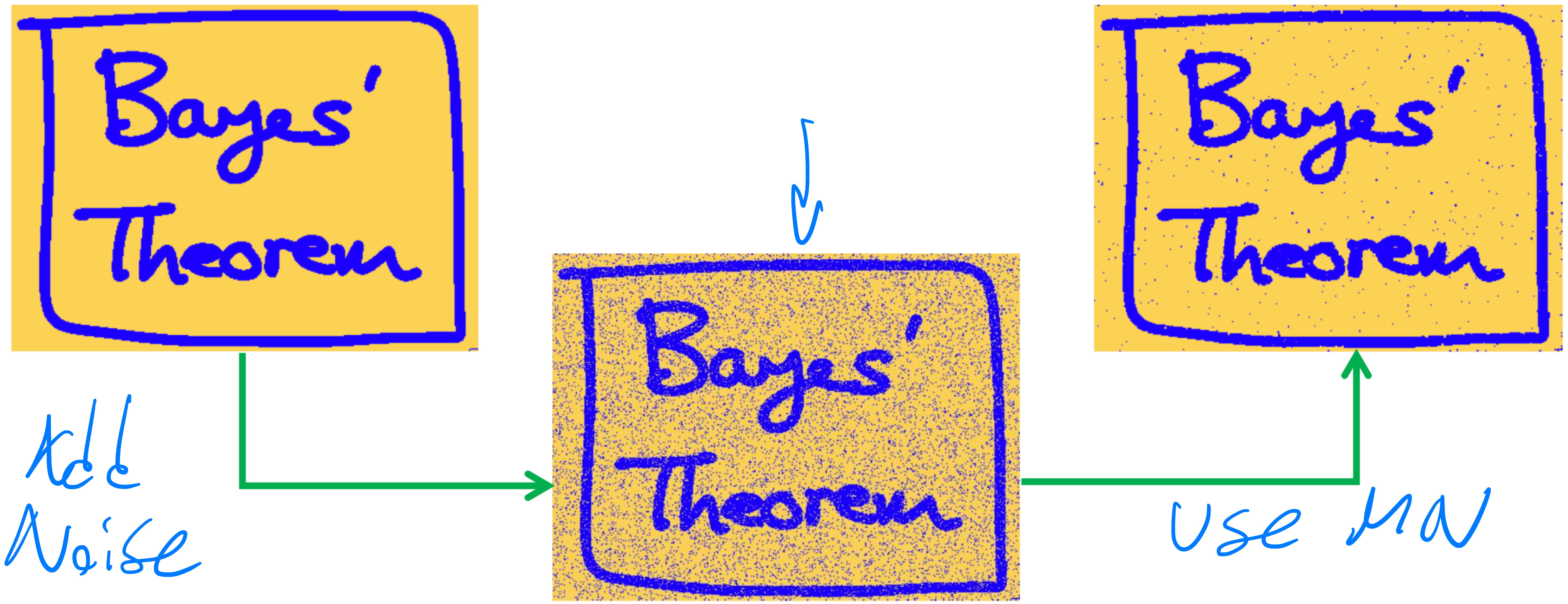
Application

Image Denoising

Christopher Burger, Christian Schuler and Stefan Harmeling, CVPR 2012

Application

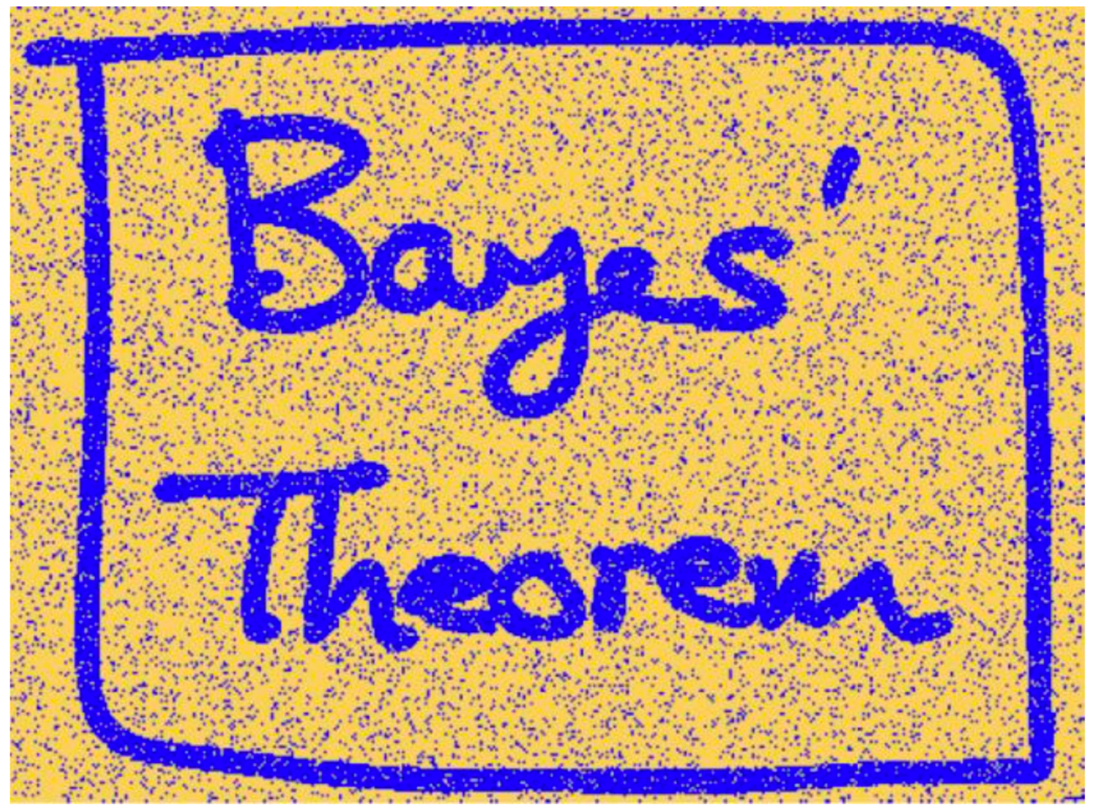
Image Denoising

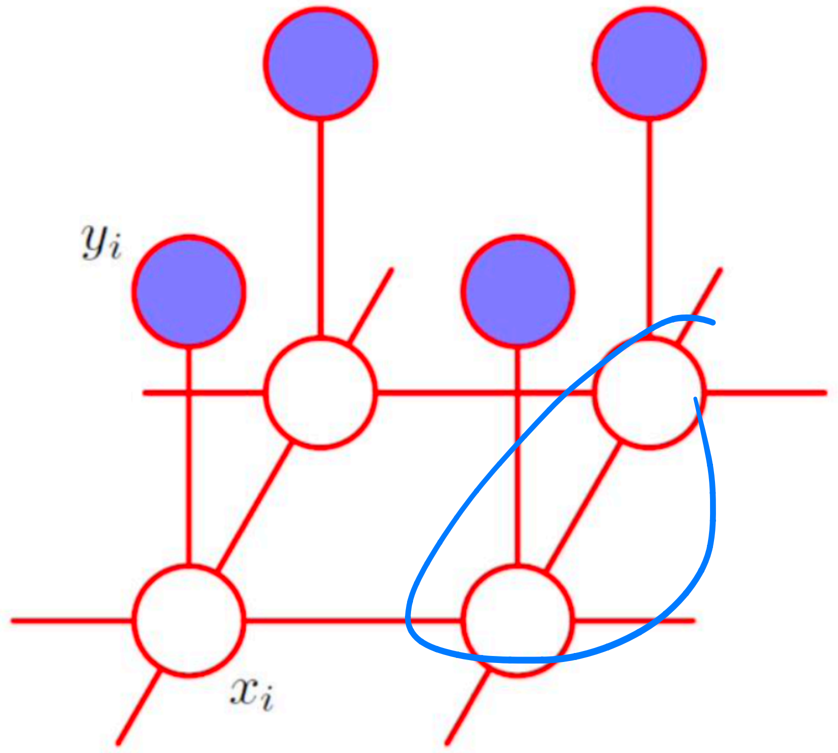


Example from Pattern Recognition and Machine Learning, Christopher Bishop

Application

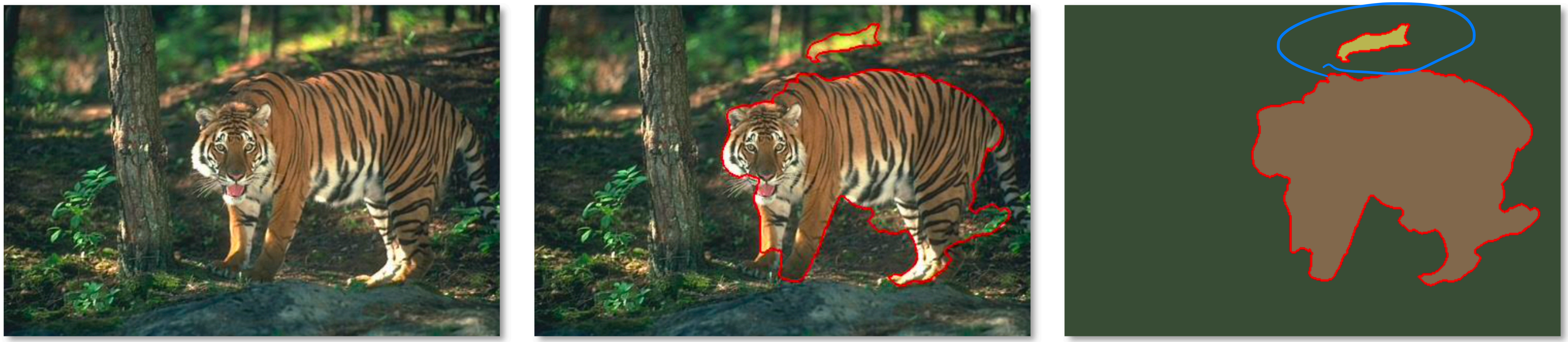
- Build a graph with each node corresponding to a pixel on the image
- Nodes are binary RVs (blue or yellow)
- x_i: initial color of pixel i
- y_i: Observed noisy color
- Adjacent pixels correspond to adjacent nodes on the graph
- Inference: $X = \arg\max_X P(X|Y)$





Application

Background Separation



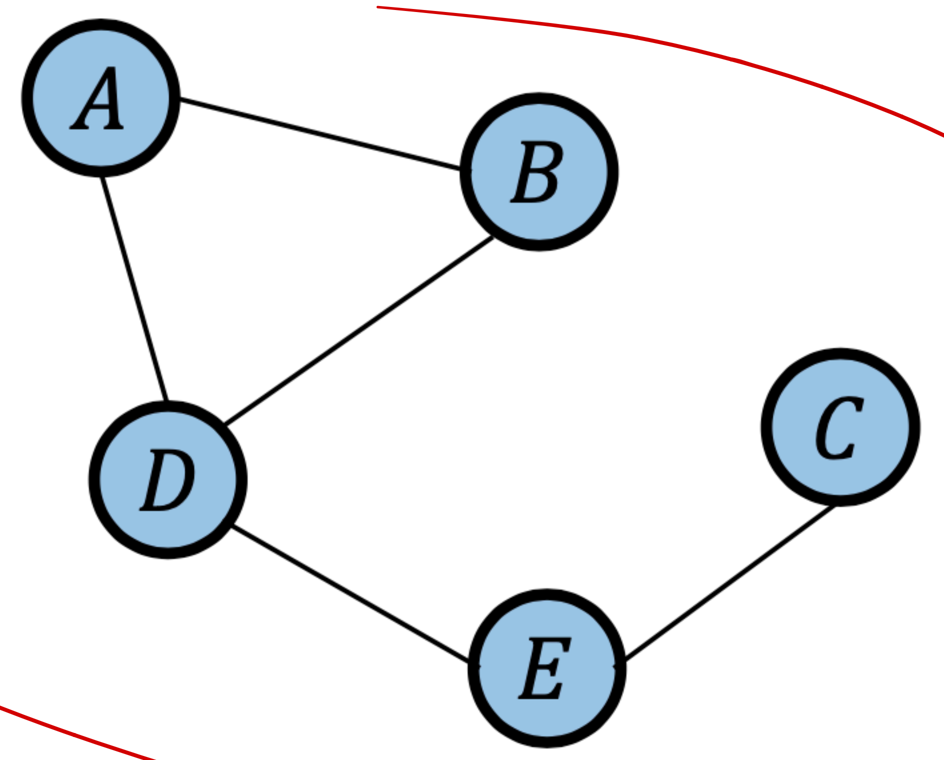
Segmentation of Multivariate Mixed Data via Lossy Coding and Compression. Yi Ma, Harm Derksen, Wei Hong and John Wright. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007.

Markov Networks

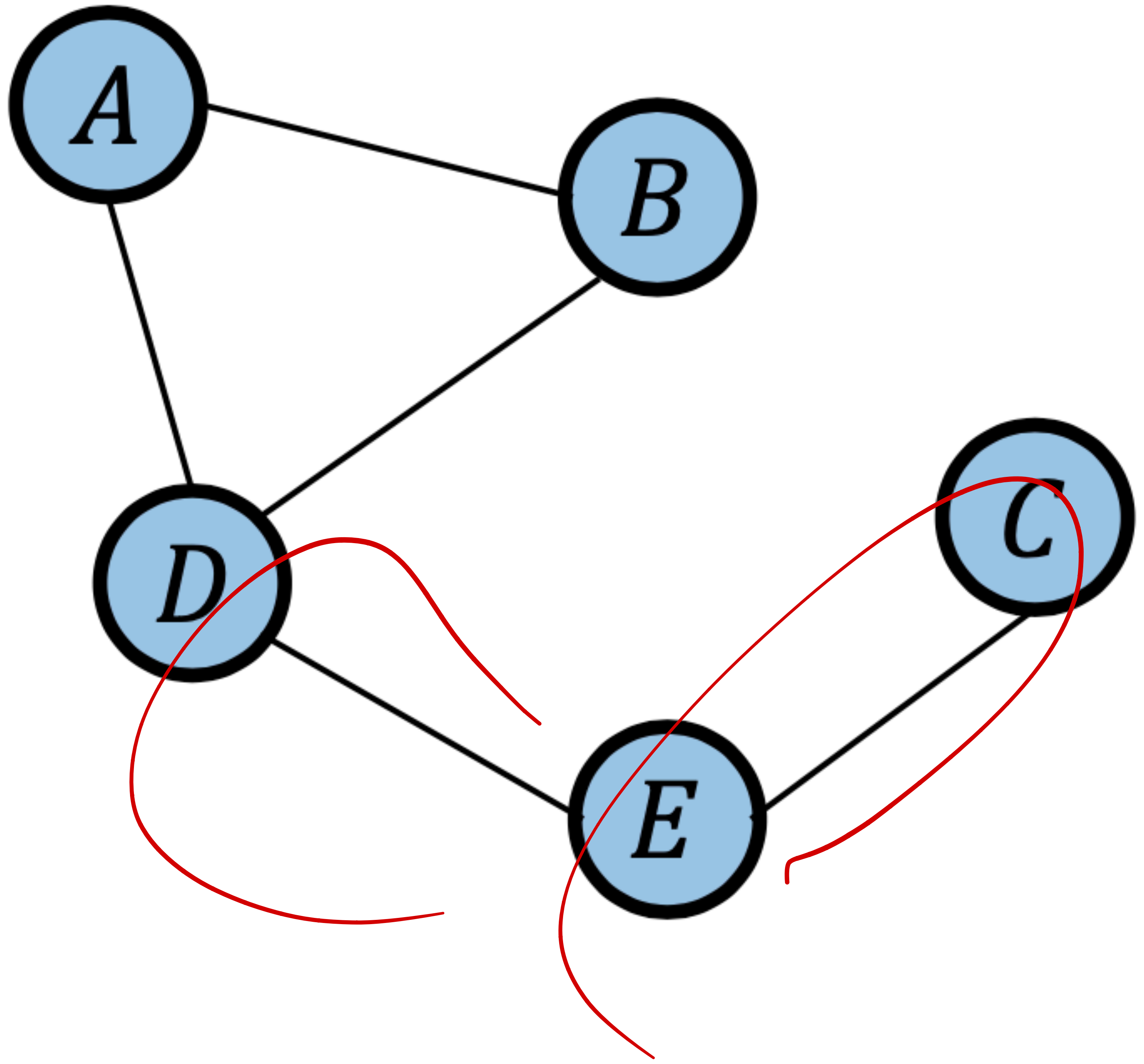
- Bayesian networks model causal relationship Smoking → cancer
- Markov networks model **correlation** or **symmetric** relationships between RVs

A, B are friends   if A likes TVshow – B TVshow

- A Markov network is a set of RVs satisfying some Markov properties described on an undirected graph

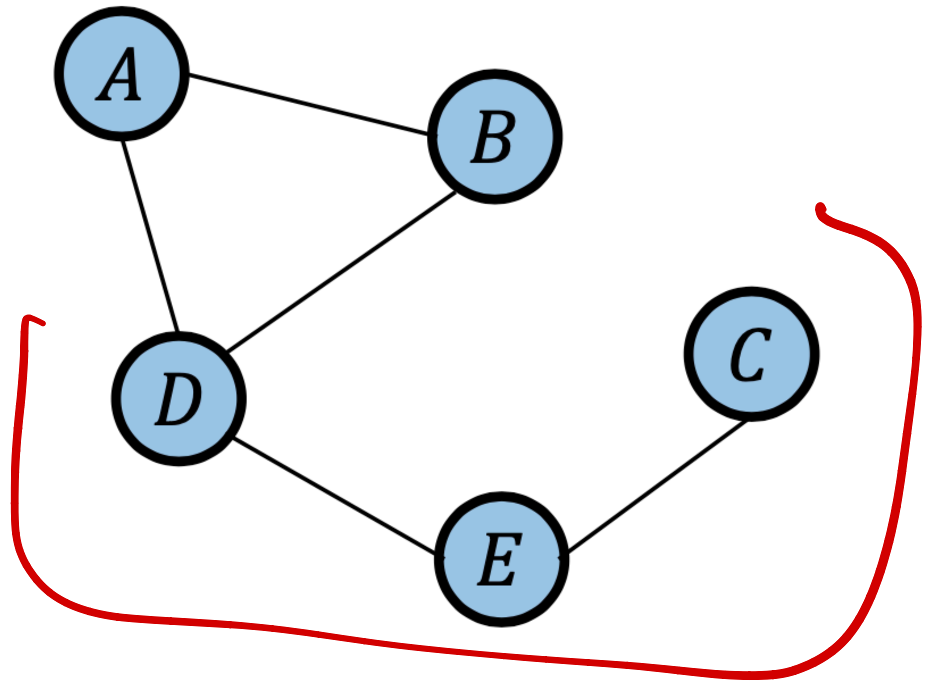


Markov Networks



Nodes — RVs
Edges — dependencies between RVs

A dep. B
A dep. D
B dep D
(...)

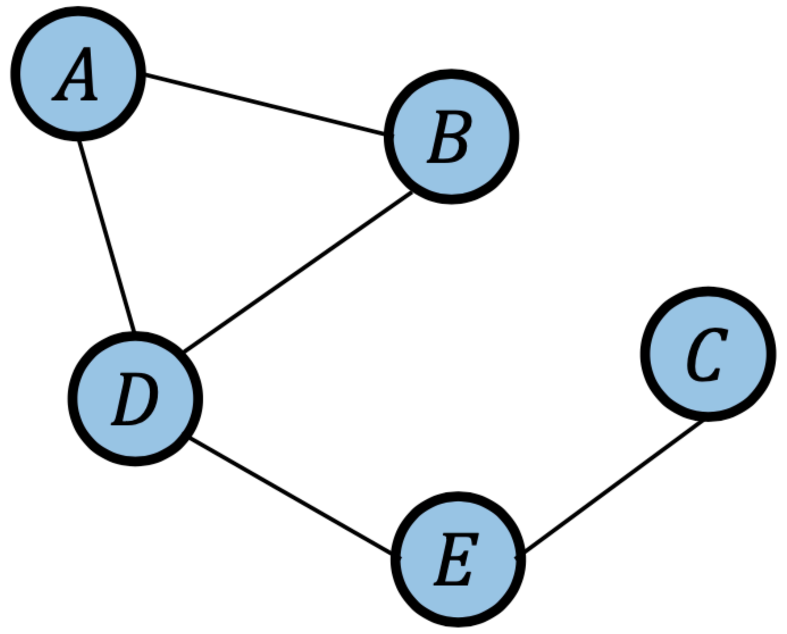


Markov Networks

Clique

- An n-clique or a complete graph with $\boldsymbol{n}$ vertices is a graph over n nodes such that there is exact one edge between any two nodes $1 \leq i \leq j \leq n$
  - 1-cliques $C_1=\{A\}, C_2=\{B\}, C_3=\{C\}, C_4=\{D\}, C_5=\{E\}$
  - 2-cliques $C_6=\{A,B\}, C_7=\{A,D\}, C_8=\{B,D\}, C_9=\{D,E\}, C_{10}=\{E,C\}$
  - 3-cliques $C_{11}=\{A,B,D\}$
  - 4-cliques ——

The set of all cliques: $\{C_1, C_2, \ldots, C_{11}\}$



Factor Potential

- Let X be a set of RVs for a clique
  - $X : \Omega \to E$
- **Factor Potentials**: Functions from E to $\mathbb{R}^+$
- **Example**: suppose X_A, X_B, X_C are binary variables, then a possible factor potential for the clique $C_{11} = \{A, B, C\}$

$$\phi([X_A = true, X_B = true, X_C = true]) = 19$$
$$\phi([X_A = true, X_B = true, X_C = false]) = 5.9$$
$$\phi([X_A = true, X_B = false, X_C = true]) = 120$$
$$\phi([X_A = true, X_B = false, X_C = flase]) = 4.6$$
$$\phi([X_A = false, X_B = true, X_C = true]) = 0.3$$
$$\phi([X_A = false, X_B = true, X_C = false]) = 2.3$$
$$\phi([X_A = false, X_B = false, X_C = true]) = 10$$
$$\phi([X_A = false, X_B = false, X_C = flase]) = 7.4$$

Definition of Markov Networks

- $X = \{X_1, \ldots, X_n\}$ - set of RVs
- $G = (X, E)$ - undirected
- $C = \{C_1, \ldots, C_m\}$ - cliques of G
- $\phi = \{\phi_1, \ldots, \phi_m\}$

Definition of Markov Networks

- $X = \{X_1, \ldots, X_n\}$
- $G = (X, E)$
- $C = \{C_1, \ldots, C_m\}$
- $\phi = \{\phi_1, \ldots, \phi_m\}$

- X is a Markov network if $P(\boxed{X_1, \ldots, X_n}) = \frac{1}{Z} \phi_1(C_1) \times \cdots \times \phi_m(C_m)$

$\propto$

Definition of Markov Networks

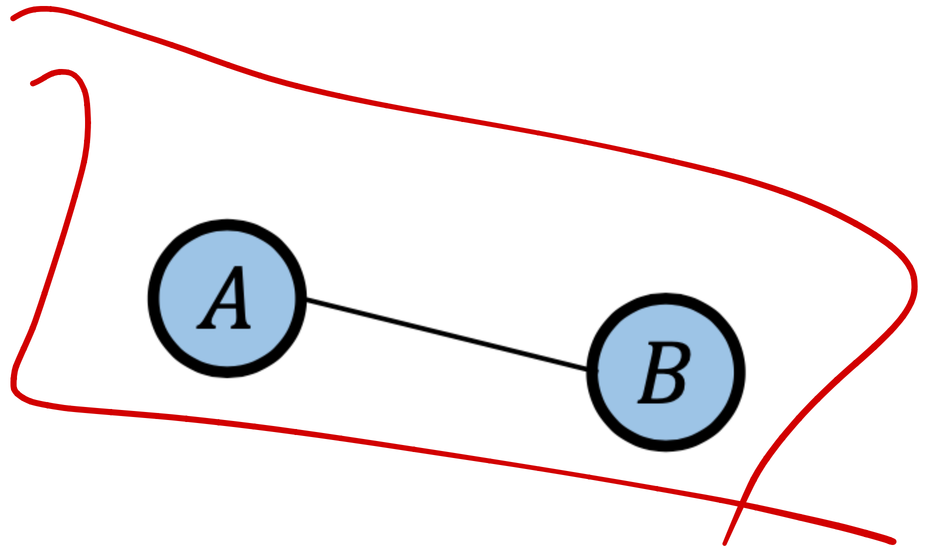
- $X = \{X_1, \ldots, X_n\}$
- $G = (X, E)$
- $C = \{C_1, \ldots, C_m\}$
- $\phi = \{\phi_1, \ldots, \phi_m\}$

- X is a Markov network if $P(X_1, \ldots, X_n) = \frac{1}{Z} \phi_1(C_1) \times \cdots \times \phi_m(C_m)$
- Z is a normalization constant $Z = \sum_{X_1,\ldots,X_n} \phi_1(C_1) \times \cdots \times \phi_m(C_m)$

Markov Networks

Example

- $\underline{X} = \{X_A, X_B\}$
- $G = (X, \{(A,B)\})$
- $C = \{C_1 = \{A\}, C_2 = \{B\}, C_3 = \{A,B\}\}$
- $\phi = \{\phi_1, \phi_2, \phi_3\}$



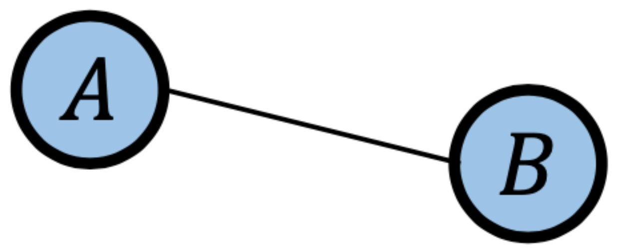
Markov Networks

Example

$$P(X_1, \dots, X_n) = \frac{1}{Z}\,\phi_1(C_1) \times \cdots \times \phi_m(C_m)$$

$$Z = \sum_{X_1,\dots,X_n} \phi_1(C_1) \times \cdots \times \phi_m(C_m)$$

- $X = \{X_A, X_B\}$
- $G = (X, \{(A,B)\})$
- $C = \{C_1 = \{A\}, C_2 = \{B\}, C_3 = \{A,B\}\}$
- $\phi = \{\phi_1, \phi_2, \phi_3\}$



$\phi_1(X_A = true) = 2, \phi_1(X_A = false) = 5$

$\phi_2(X_B = true) = 3, \phi_2(X_B = false) = 1$

$\phi_3(X_A = true, X_B = true) = 13, \dots$

$$Z = \sum_{X_A X_B} \phi_1(X_A) \cdot \phi_2(X_B) \cdot \phi_3(X_A, X_B)$$

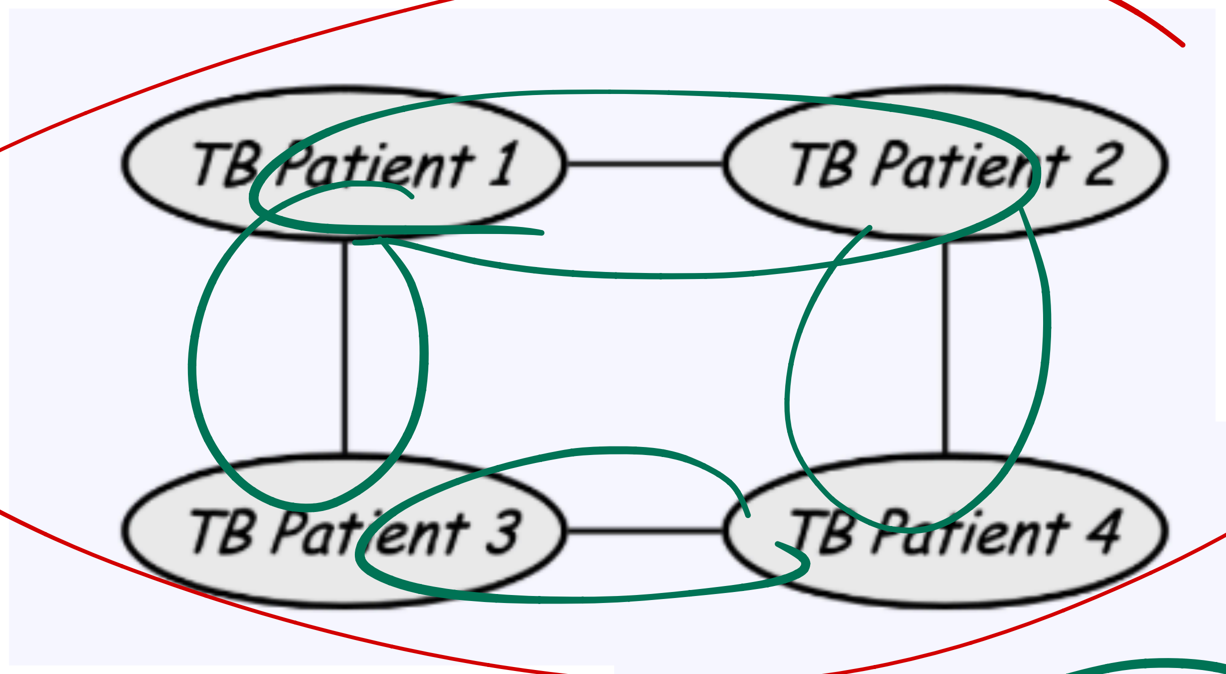
$$P(X_A = true, X_B = false) = \frac{1}{Z}\phi_1(X_A = true)\,\phi_2(X_B = false)\,\phi_3(X_A = true, X_B = false)$$

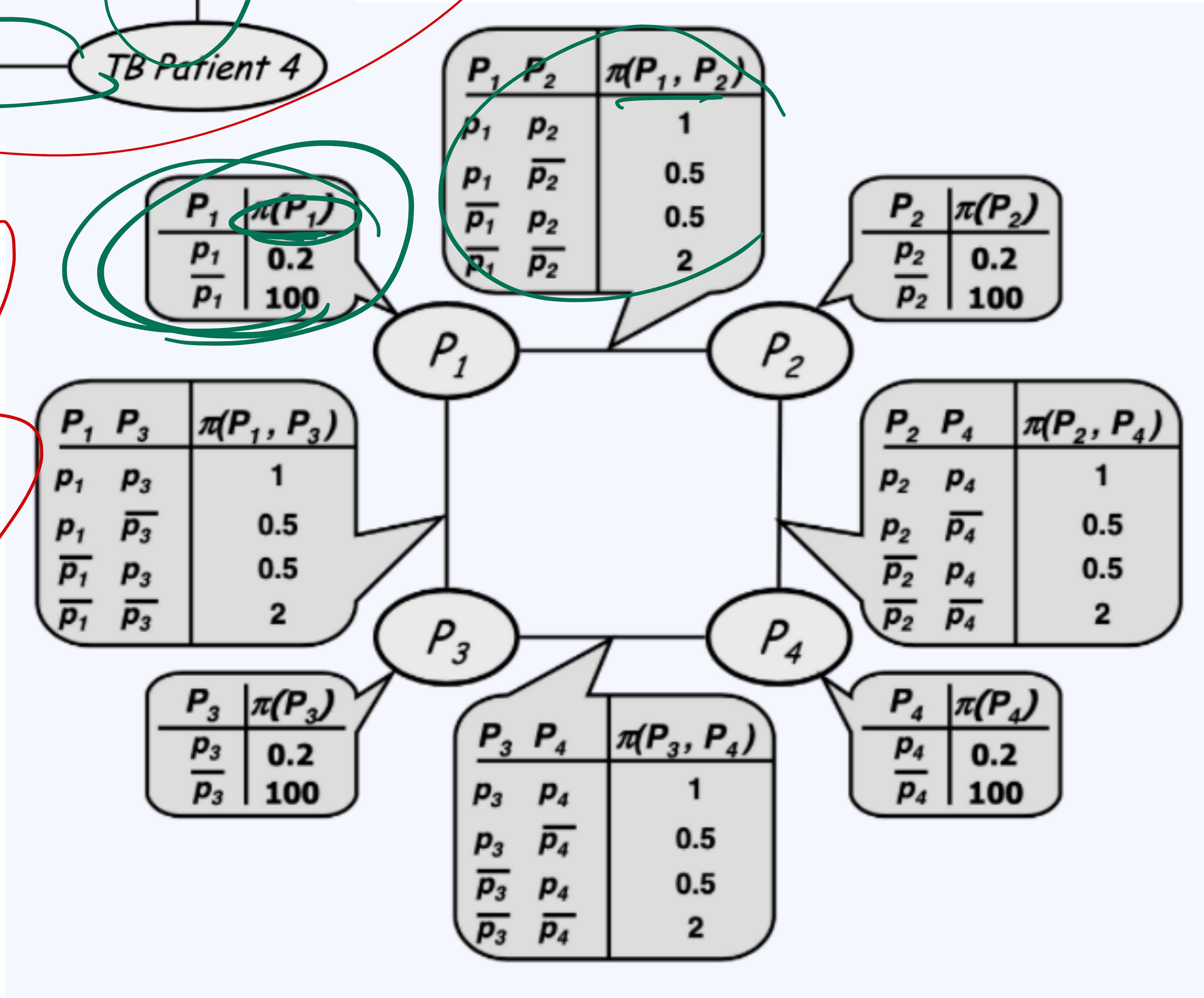
Markov Networks

Example

- $X = \{P_1, P_2, P_3, P_4\}$
- $G = (X, \{(P_1, P_2), (P_2, P_4), (P_3, P_4), (P_1, P_3)\})$

$\phi_m = \pi_i$





P_1	$\pi(P_1)$
p_1	0.2
$\bar{p}_1$	100

P_1	P_2	$\pi(P_1, P_2)$
p_1	p_2	1
p_1	$\bar{p}_2$	0.5
$\bar{p}_1$	p_2	0.5
$\bar{p}_1$	$\bar{p}_2$	2

P_2	$\pi(P_2)$
p_2	0.2
$\bar{p}_2$	100

P_1	P_3	$\pi(P_1, P_3)$
p_1	p_3	1
p_1	$\bar{p}_3$	0.5
$\bar{p}_1$	p_3	0.5
$\bar{p}_1$	$\bar{p}_3$	2

P_2	P_4	$\pi(P_2, P_4)$
p_2	p_4	1
p_2	$\bar{p}_4$	0.5
$\bar{p}_2$	p_4	0.5
$\bar{p}_2$	$\bar{p}_4$	2

P_3	$\pi(P_3)$
p_3	0.2
$\bar{p}_3$	100

P_3	P_4	$\pi(P_3, P_4)$
p_3	p_4	1
p_3	$\bar{p}_4$	0.5
$\bar{p}_3$	p_4	0.5
$\bar{p}_3$	$\bar{p}_4$	2

P_4	$\pi(P_4)$
p_4	0.2
$\bar{p}_4$	100

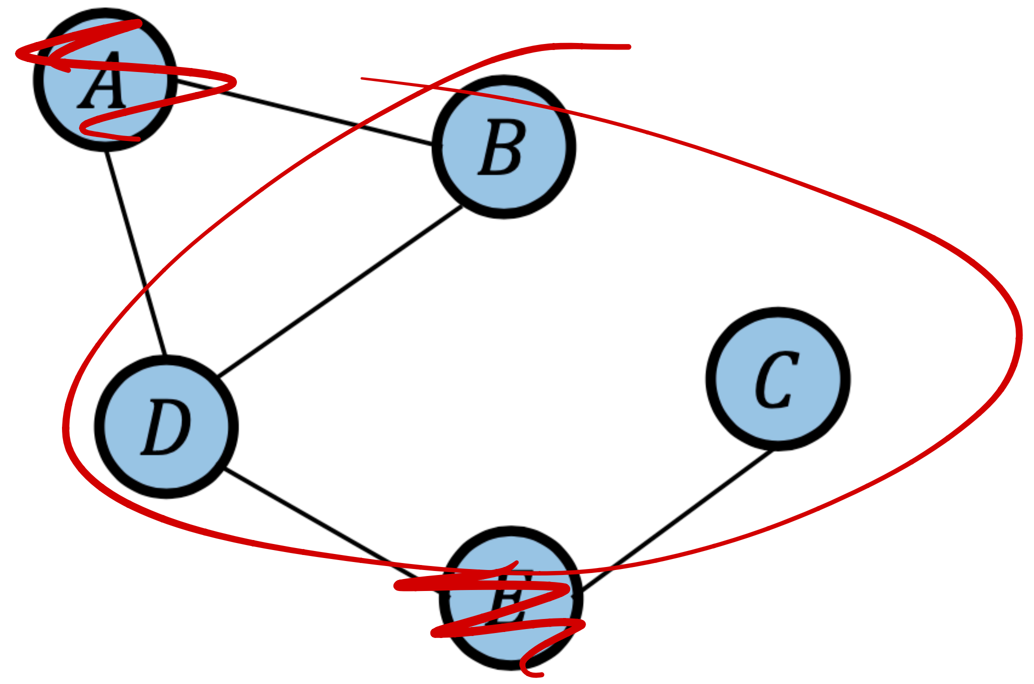
Independence

- **X** is independent of **Y** given **Z**: $(X \perp Y) | Z$
  - $P(X, Y | Z) = \boxed{P(X | Z)P(Y | Z)}$
- **Pairwise Markov property:**
  - Any two non-adjacent RVs X_i, X_j are conditionally independent given all other RVs in X

$(X_i \perp X_j) \,|\, (X - \{X_i, X_j\})$

E.g. $(A \perp E) \,|\, (B, C, D)$

$(B \perp D) \,|\, (A, C, E)$

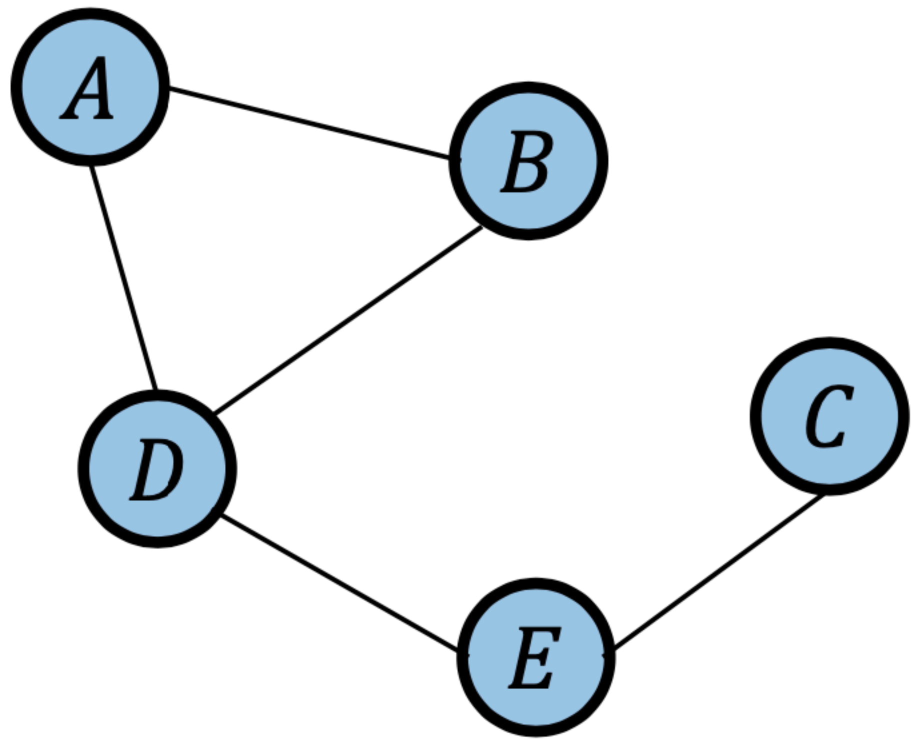


Local Markov Property

- **Local Markov property**: a variable is conditionally independent of all other variables given its neighbors.
- Let $N(X_i)$ be all neighbors of X_i, then $(X_i \perp \{X - \{X_i, N(X_i)\}\}) | N(X_i)$

$(D \perp C) | (A, B, E)$

$(B \perp \{C, E\}) | (A, D)$



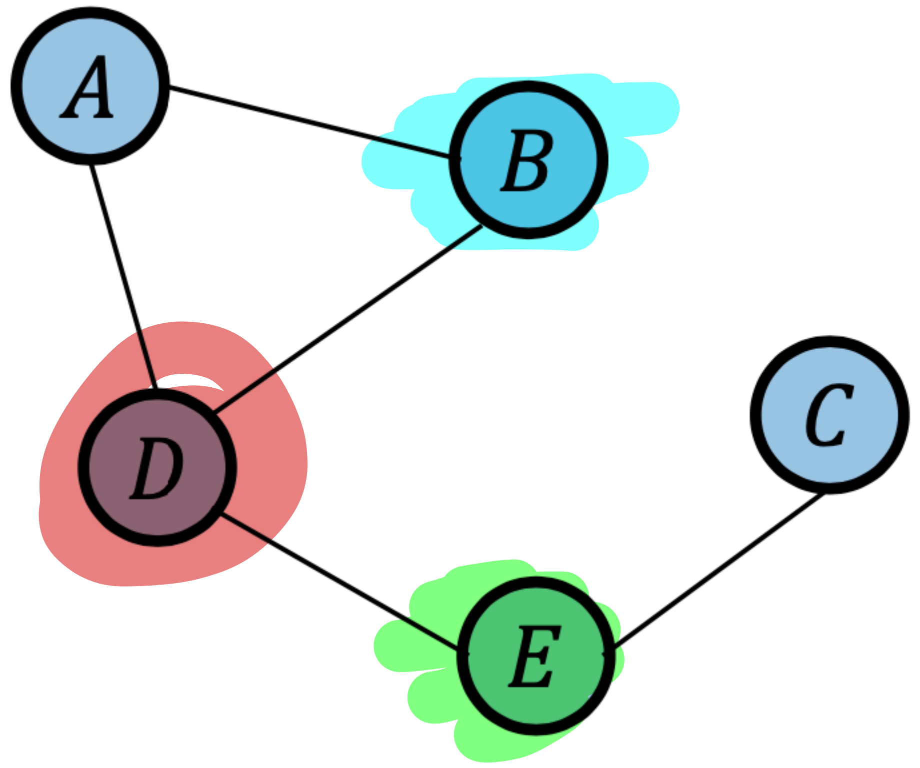
Global Markov Property

- **Global Markov property**: any two subsets X^1, X^2 are conditionally independent given a separating set $S \subset X$.

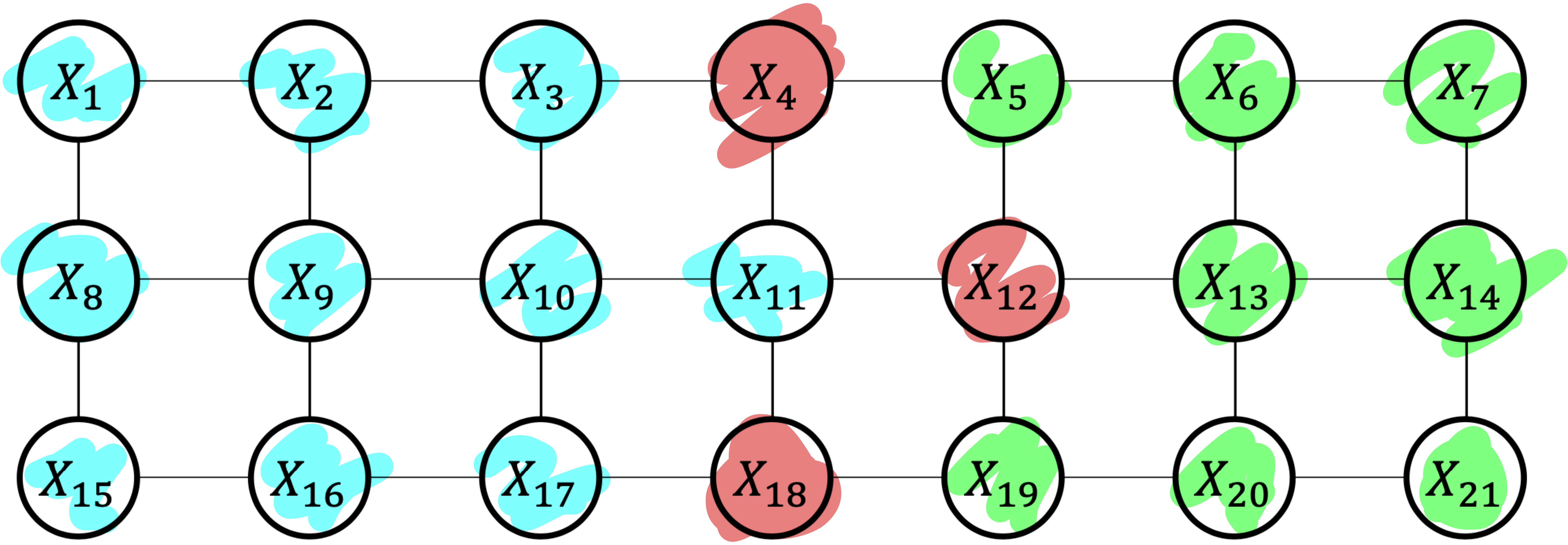
$$(X^1 \perp X^2)|S$$

$(A \perp C)|(D)$

$(B \perp E)|(D)$



2D Grid Example



$S = \{X_4, X_{12}, X_{18}\}$

$A = \{X_1, \ldots, X_3, X_8, \ldots, X_{11}, X_{15}, \ldots, X_{17}\}$

$B = \{X_5, \ldots, X_7, X_{13}, X_{14}, X_{19}, \ldots, X_{21}\}$

$$P(X_8, X_{14} \mid X_4, X_{12}, X_{18}) = P(X_8 \mid X_4, X_{12}, X_{18})\, P(X_{14} \mid X_4, X_{12}, X_{18})$$

Hammersley-Clifford Theorem

- Given RV set X, graph $G = (X, E)$, if X satisfies the global Markov property, then

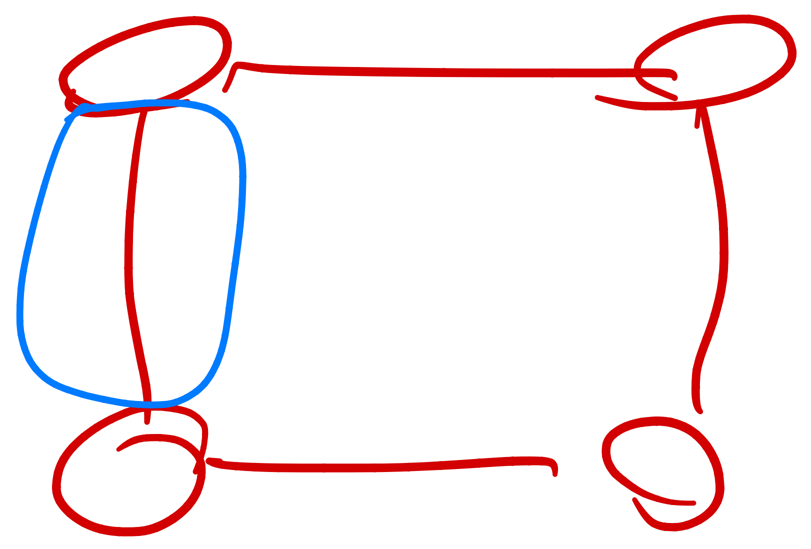
$$P(X_1, \ldots, X_n) = \frac{1}{Z}\phi(C_1) \times \ldots \times \phi_m(C_m)$$

Factor Potentials

- Suppose X_1 represents *Patient 1 has TB*
  - Make $\phi_1(C_1 = \{X_1\})$ useful:
    - $f_1[1]$:Weight decreasing?
    - $f_1[2]$: Fever?
    - $f_1[3]$:X-ray positive?
- Use weights:

$$\phi_1(C_1) = \exp\left(w_{C_1}[1]f_1[1] + w_{C_1}[2]f_1[2] + w_{C_1}[3]f_1[3]\right)$$

Factor Potentials



- Use weights:

$$\phi_1(C_1) = \exp\left(w_{C_1}[1]f_1[1] + w_{C_1}[2]f_1[2] + w_{C_1}[3]f_1[3]\right)$$

- In general

$$\phi_i(C_i) = \exp\left(\sum_{j=1}^{k} w_{C_i}[j]f_i[j]\right)$$

$f_i[j]$ - describes properties of an edge

Logistic Model

$$P(X_1, \dots, X_n) = \frac{1}{Z}\prod_{i=1}^{m}\phi_i(C_i) = \frac{1}{Z}\exp\left(\sum_{i=1}^{m}\sum_{j=1}^{k} w_{C_i}[j]f_i[j]\right)$$

- Usefulness?

$$\log P(X_1, \dots, X_n) = \log\frac{1}{Z} + \sum_{i=1}^{m}\sum_{j=1}^{k} w_{C_i}[j]\,f_i[j]$$

Associative Markov Networks

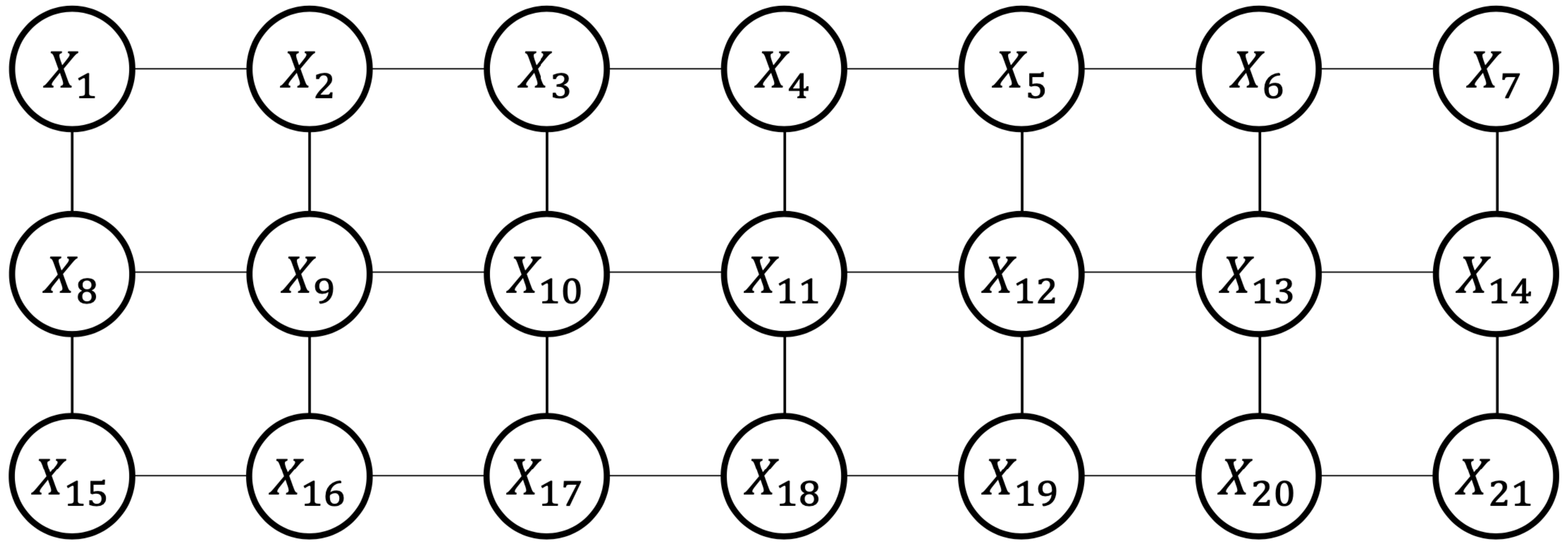
- Associative Markov networks are a special variant of pairwise Markov networks that proves successful in computer vision applications
- In pairwise Markov networks, there are only two types of cliques

- Factor potential:

$\phi_{node}(X_i) = \exp(\sum_k sign(X_i) w_{node}[k] f_i[k])$

$\phi_{edge}(X_i, X_j) = \exp(\sum_k w_{edge}[k] f_{i,j}[k])$

$sign(X_i) = 1$ if $X_i = true$ and $sign(X_i) = -1$ if $X_i = false$



Associative Markov Networks

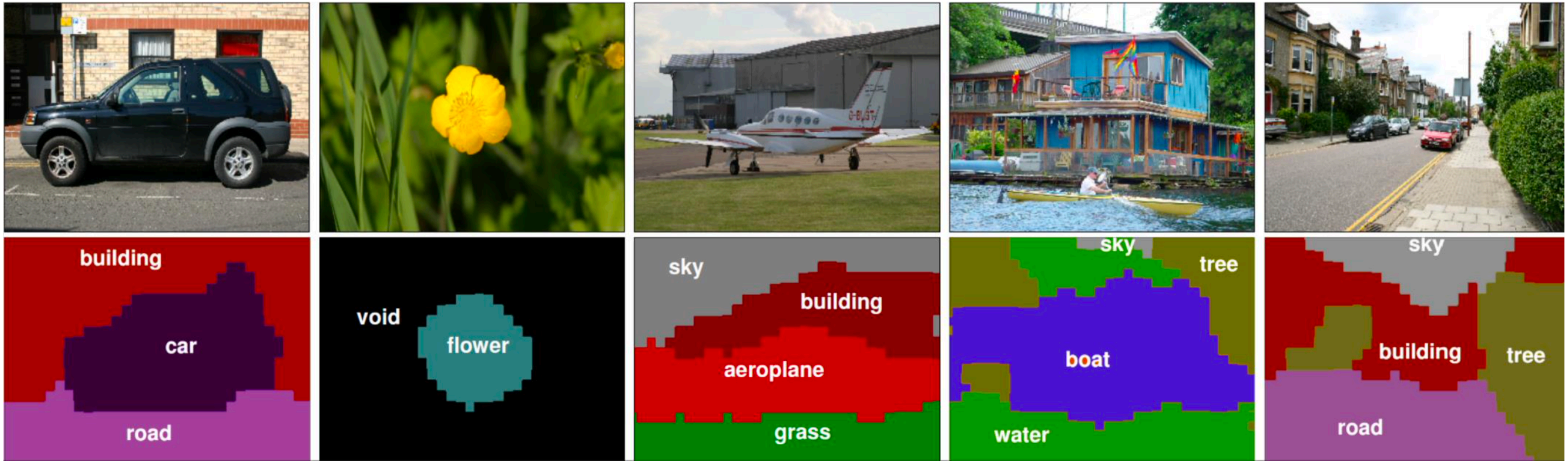
- In associative Markov networks, the edge potentials are further constrained as

$$\phi_{edge}(X_i, X_j) = \exp\left(\sum_k w_{edge}[k] f_{i,j}[k]\right), \quad if\ X_i = X_j$$
$$\phi_{edge}(X_i, X_j) = \exp(0) = 1, \quad if\ X_i \neq X_j$$

- Joint Probability Distribution

$$P(X) = \frac{1}{Z} \exp\left(\sum_{X_i} \sum_k sign(X_i) w_{node}[k] f_i[k] + \sum_{\substack{(X_i, X_j)\ s.t.\\ X_i = X_j\ and\\ X_i, X_j\ adjacent}} \sum_k w_{edge}[k] f_{i,j}[k] \right)$$

Image Segmentation



Jakob Verbeek, Bill Triggs. Region Classication with Markov Field Aspect Models. In CVPR 2007

Image Segmentation

Flower

- Overlay an associate Markov network

$X_i \sim$ Patch of Pixels

x_i = true ⇒ patch i is part of Flower

x_i = false ⇒ patch i is NOT Part of Flower

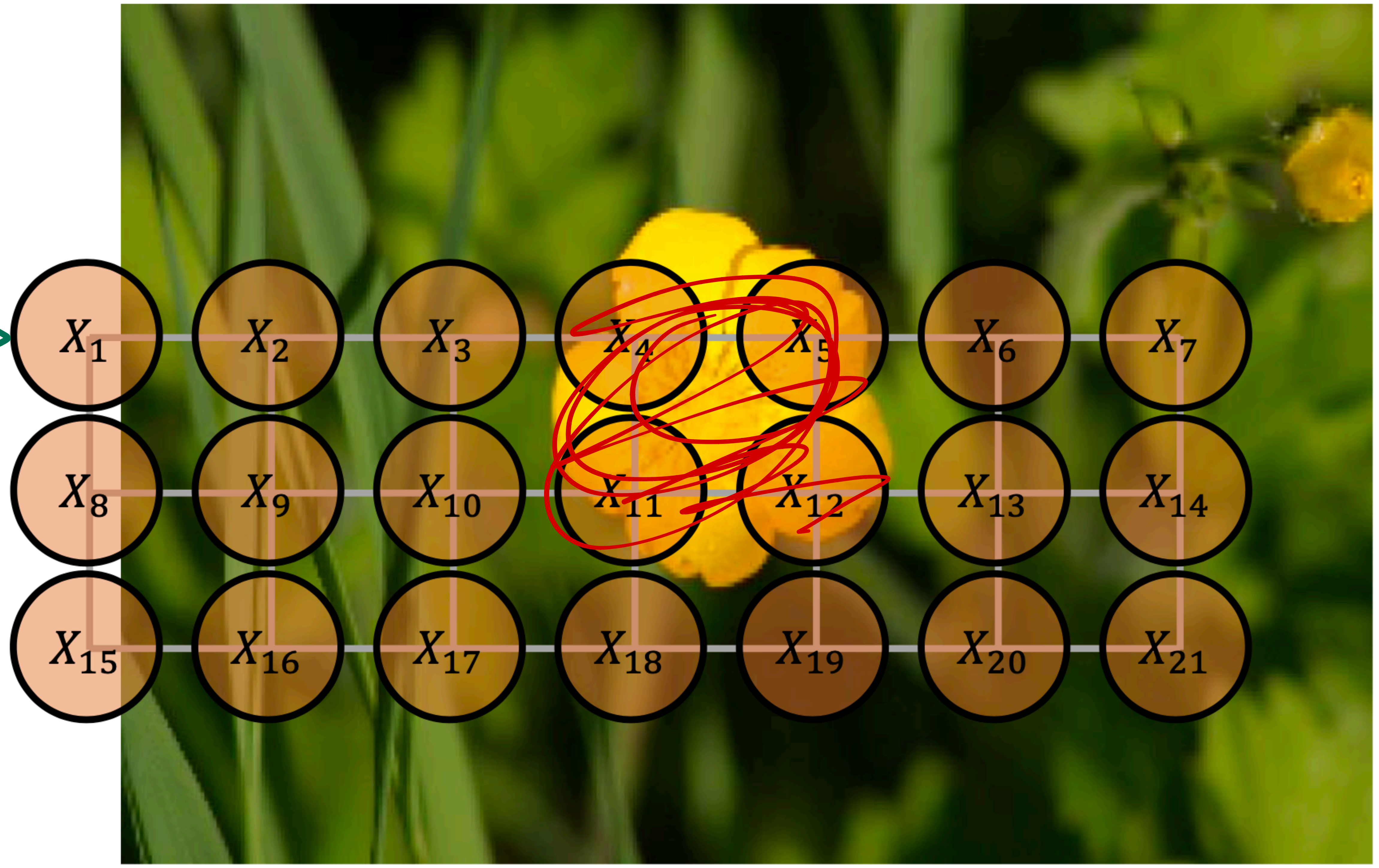


Image Segmentation

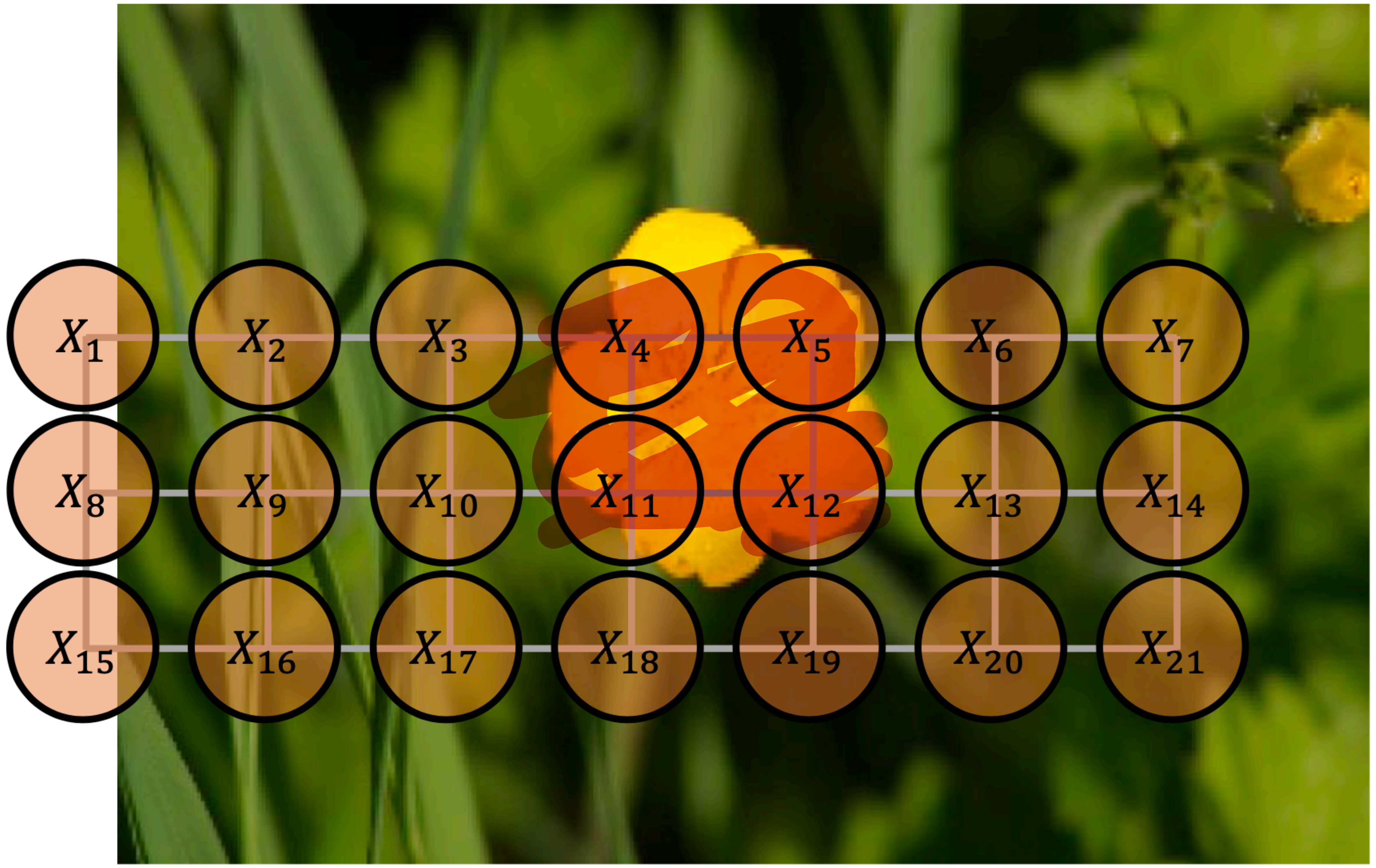
Flower

- Joint Probability

$$P(X) = \frac{1}{Z} \exp\left(\sum_{X_i} \sum_{k} sign(X_i) w_{node}[k] f_i[k] + \sum_{\substack{(X_i,X_j)\ s.t.\\ X_i = X_j\ and\\ X_i,X_j\ adjacent}} \sum_{k} w_{edge}[k] f_{i,j}[k] \right)$$

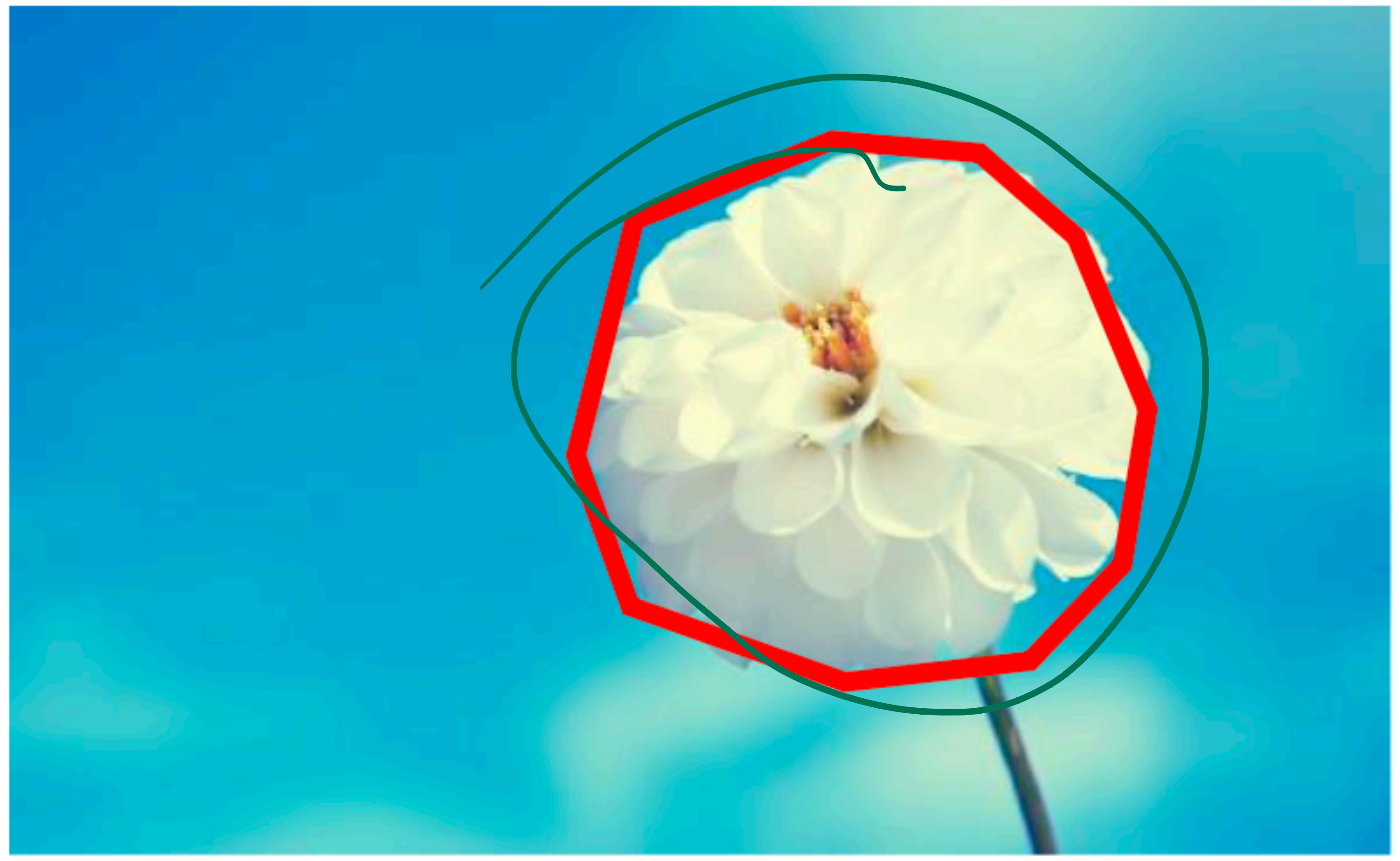
P(x) – highest when x_4, x_5, x_{11}, x_{12} true and rest false

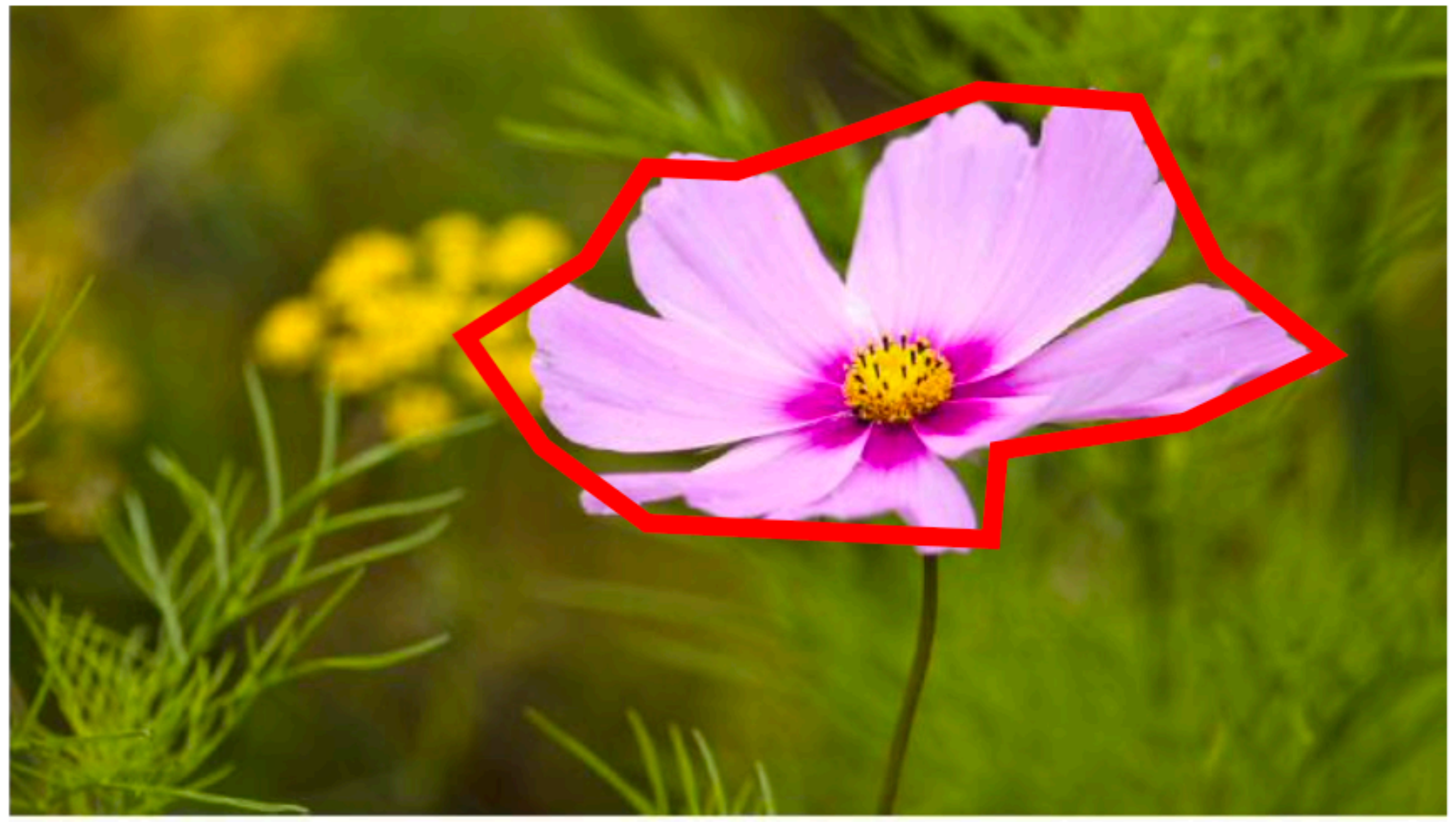
- Training:
  - Pick features $f_i, f_{i,j}$
  - Learn w_{node} and w_{edge}



Learning w_{node} and w_{edge}

- Use segmented images





Learning w_{node} and w_{edge}

- Use segmented images

Compute W_{node} W_{edges} that maximizes $P(X)$