

# Graphical Models

## Inference and Learning

### Lecture 1

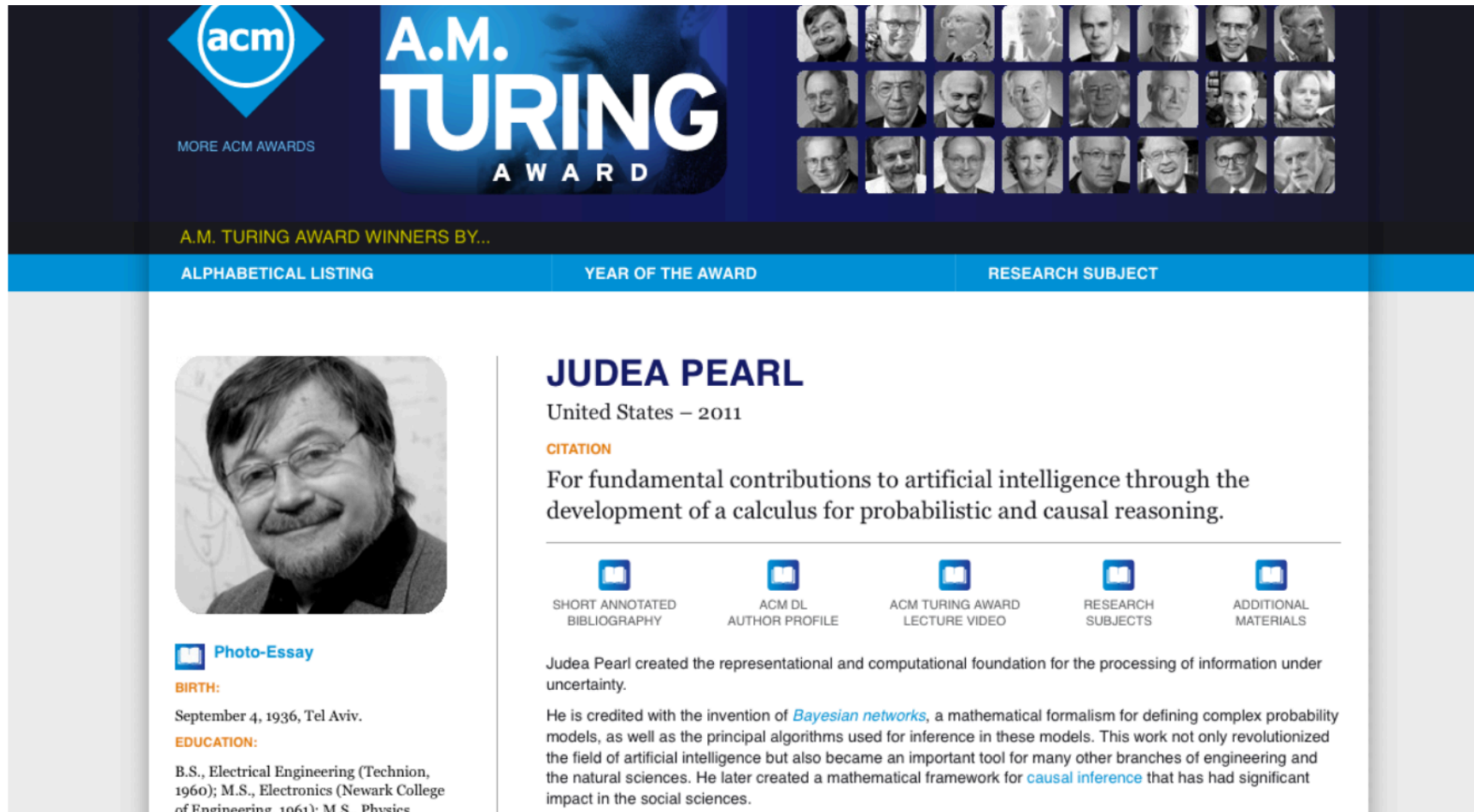
MVA

2018 – 2019

<http://thoth.inrialpes.fr/~alahari/disinflern>

Slides based on material from Stephen Gould, Pushmeet Kohli, Nikos Komodakis, M. Pawan Kumar, Carsten Rother, Daphne Koller, Dhruv Batra

# Graphical Models ?



The screenshot shows the ACM Turing Award website. At the top, there is a navigation bar with the ACM logo, the text 'A.M. TURING AWARD', and a grid of 24 small portraits of award winners. Below this is a menu with three options: 'ALPHABETICAL LISTING', 'YEAR OF THE AWARD', and 'RESEARCH SUBJECT'. The main content area features a large portrait of Judea Pearl on the left. To the right of the portrait, his name 'JUDEA PEARL' is displayed in large blue letters, followed by 'United States – 2011'. Below this is a 'CITATION' section with the text: 'For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.' Underneath the citation are five icons representing different resources: 'SHORT ANNOTATED BIBLIOGRAPHY', 'ACM DL AUTHOR PROFILE', 'ACM TURING AWARD LECTURE VIDEO', 'RESEARCH SUBJECTS', and 'ADDITIONAL MATERIALS'. A paragraph of text describes his work on Bayesian networks and causal inference. At the bottom left, there is a 'Photo-Essay' link and a 'BIRTH:' section with the date 'September 4, 1936, Tel Aviv.' and an 'EDUCATION:' section listing his degrees from Technion, Newark College of Engineering, and a Physics M.S.

**acm**  
MORE ACM AWARDS

**A.M. TURING AWARD**

A.M. TURING AWARD WINNERS BY...

ALPHABETICAL LISTING    YEAR OF THE AWARD    RESEARCH SUBJECT

**JUDEA PEARL**  
United States – 2011

**CITATION**  
For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

SHORT ANNOTATED BIBLIOGRAPHY    ACM DL AUTHOR PROFILE    ACM TURING AWARD LECTURE VIDEO    RESEARCH SUBJECTS    ADDITIONAL MATERIALS

Judea Pearl created the representational and computational foundation for the processing of information under uncertainty.

He is credited with the invention of *Bayesian networks*, a mathematical formalism for defining complex probability models, as well as the principal algorithms used for inference in these models. This work not only revolutionized the field of artificial intelligence but also became an important tool for many other branches of engineering and the natural sciences. He later created a mathematical framework for *causal inference* that has had significant impact in the social sciences.

**Photo-Essay**

**BIRTH:**  
September 4, 1936, Tel Aviv.

**EDUCATION:**  
B.S., Electrical Engineering (Technion, 1960); M.S., Electronics (Newark College of Engineering, 1961); M.S., Physics

# What this class is about?

- Making **global** predictions from **local** observations

**Inference**

- Learning such models from large quantities of data

**Learning**

# Motivation

- Consider the example of medical diagnosis



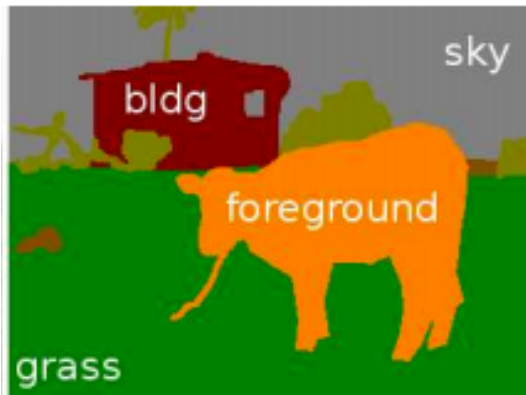
Predisposing factors  
Symptoms  
Test results



Diseases  
Treatment outcomes

# Motivation

- A very different example: image segmentation



Millions of pixels  
Colours / features



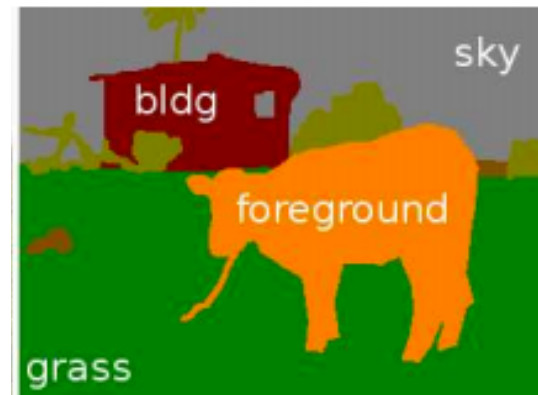
Pixel labels  
{building, grass, cow, sky}

e.g., [He et al., 2004; Shotton et al., 2006; Gould et al., 2009]

Slide inspired by PGM course, Daphne Koller

# Motivation

- What do these two problems have in common?



Slide inspired by PGM course, Daphne Koller

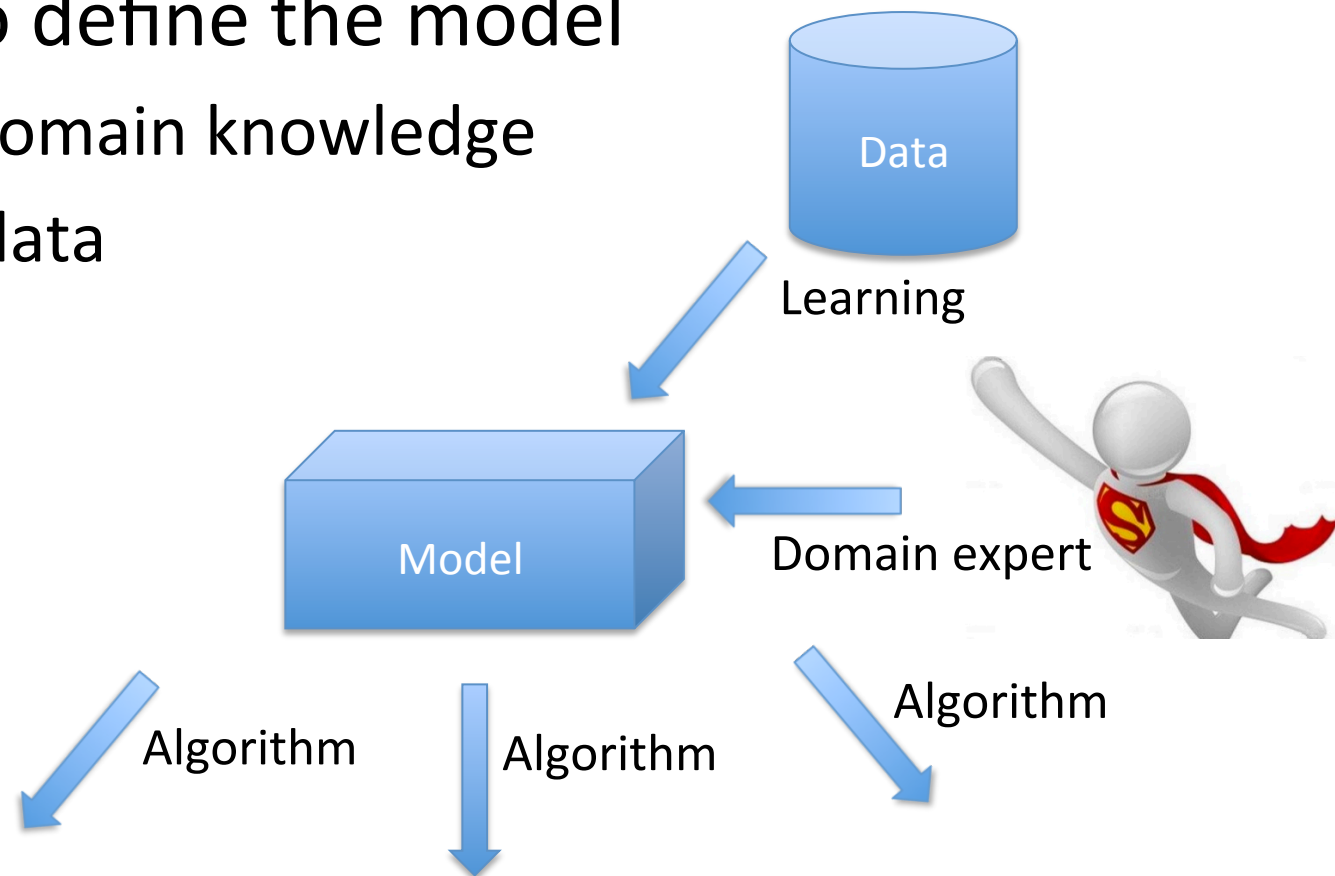
# Motivation

- What do these two problems have in common?
  - Many variables
  - Uncertainty about the correct answer

Graphical Models (or Probabilistic Graphical Models)  
provide a framework to address these problems

# (Probabilistic) Graphical Models

- First, it is a model: a declarative representation
- Can also define the model
  - with domain knowledge
  - from data



Slide inspired by PGM course, Daphne Koller



# (Probabilistic) Graphical Models

- Why probabilistic ?
- To model uncertainty
- Uncertainty due to:
  - Partial knowledge of state of the world
  - Noisy observations
  - Phenomena not observed by the model
  - Inherent stochasticity

# (Probabilistic) Graphical Models

- Probability theory provides
  - Standalone representation with clear semantics
  - Reasoning patterns (conditioning, decision making)
  - Learning methods

# (Probabilistic) Graphical Models

- Why graphical ?
- Intersection of ideas from probability theory and computer science
  - To represent large number of variables

Predisposing factors

Symptoms

Test results

Millions of pixels

Colours / features

**Random variables  $Y_1, \dots, Y_n$**

Goal: capture uncertainty through joint distribution  $P(Y_1, \dots, Y_n)$

# (Probabilistic) Graphical Models

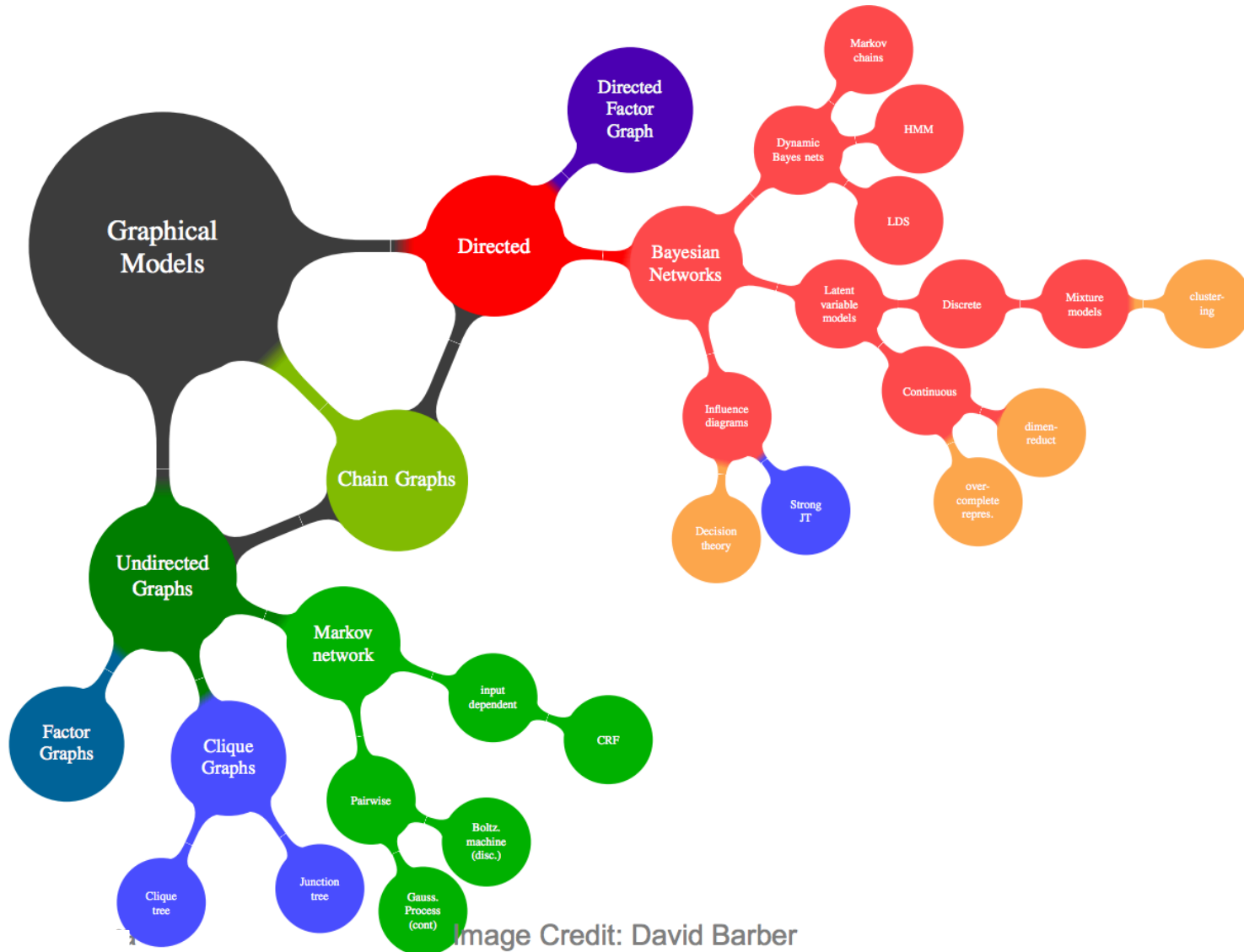
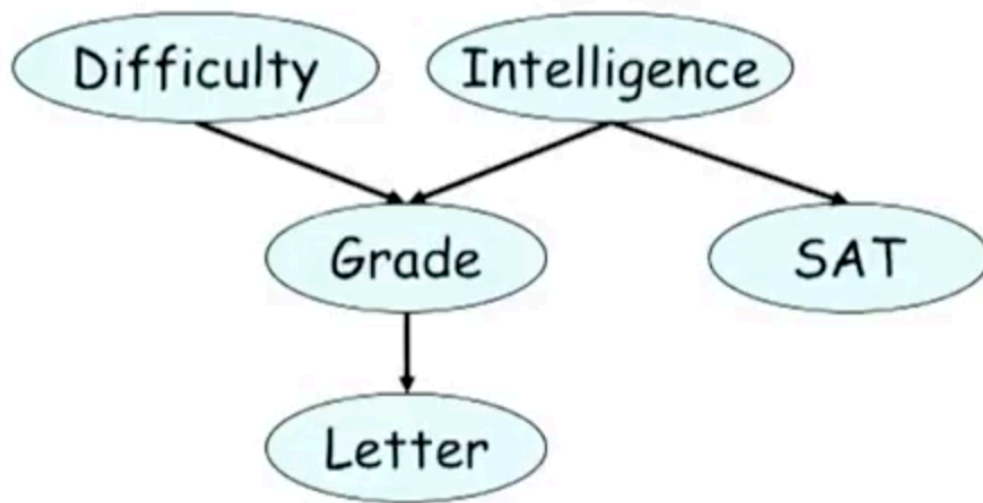


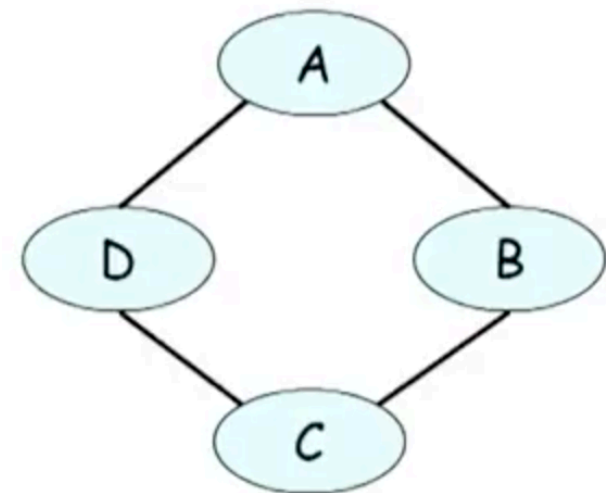
Image Credit: David Barber

# (Probabilistic) Graphical Model

- Examples



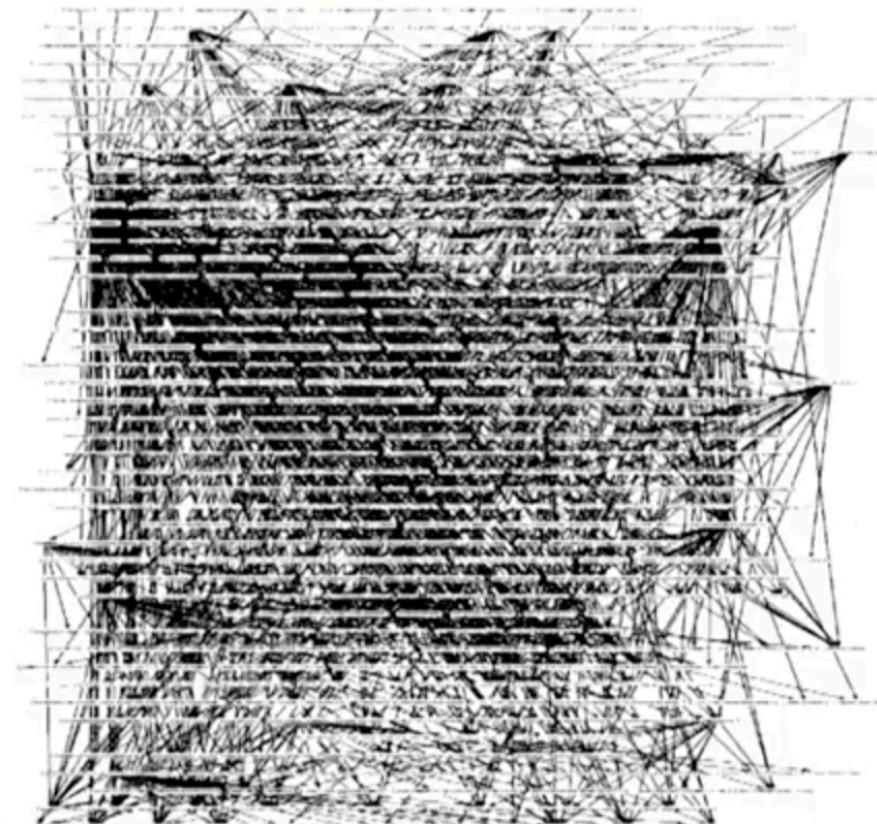
Bayesian network  
(directed graph)



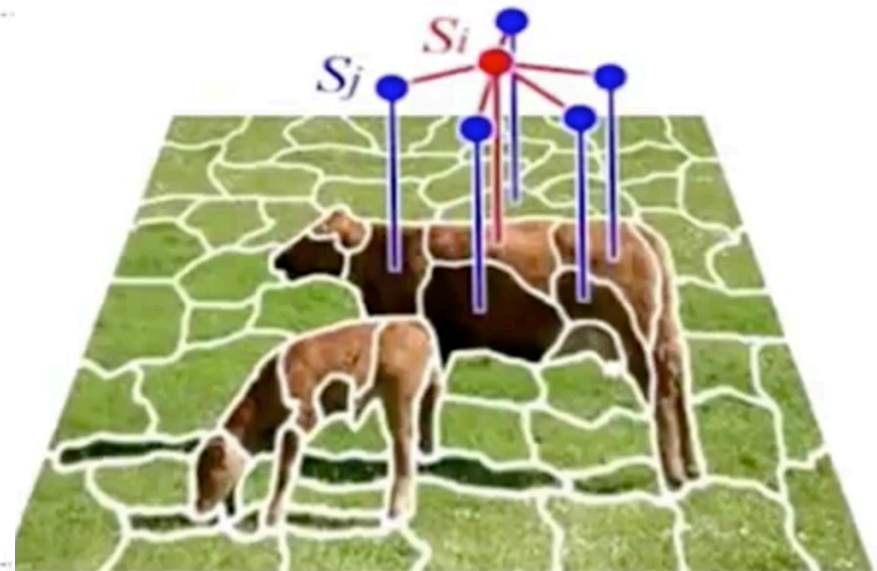
Markov network  
(undirected graph)

# (Probabilistic) Graphical Model

- Examples



Diagnosis network: Pradhan et al., UAI'94



Segmentation network (Courtesy D. Koller)

# (Probabilistic) Graphical Model

- Intuitive & compact data structure
- Efficient reasoning through general-purpose algorithms
- Sparse parameterization
  - Through expert knowledge, or
  - Learning from data

# (Probabilistic) Graphical Model

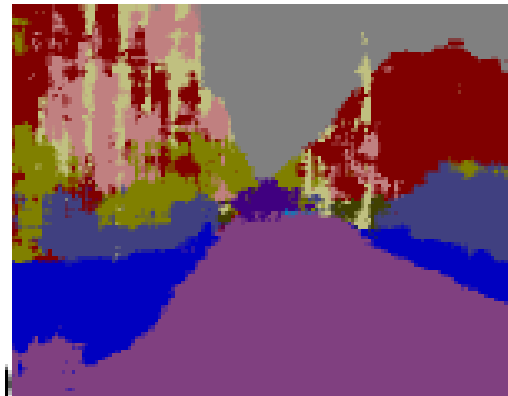
- Many many applications
  - Medical diagnosis
  - Fault diagnosis
  - Natural language processing
  - Traffic analysis
  - Social network models
  - Message decoding
  - Computer vision: segmentation, 3D, pose estimation
  - Speech recognition
  - Robot localization & mapping



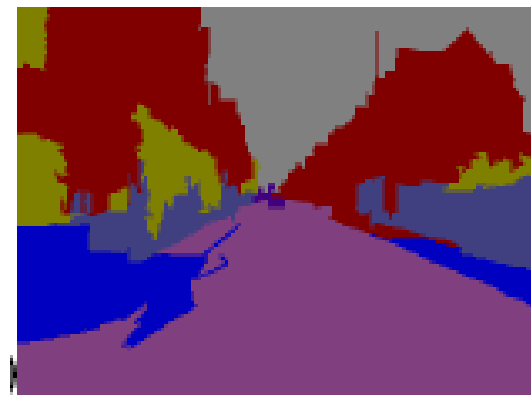
# Image segmentation



Image



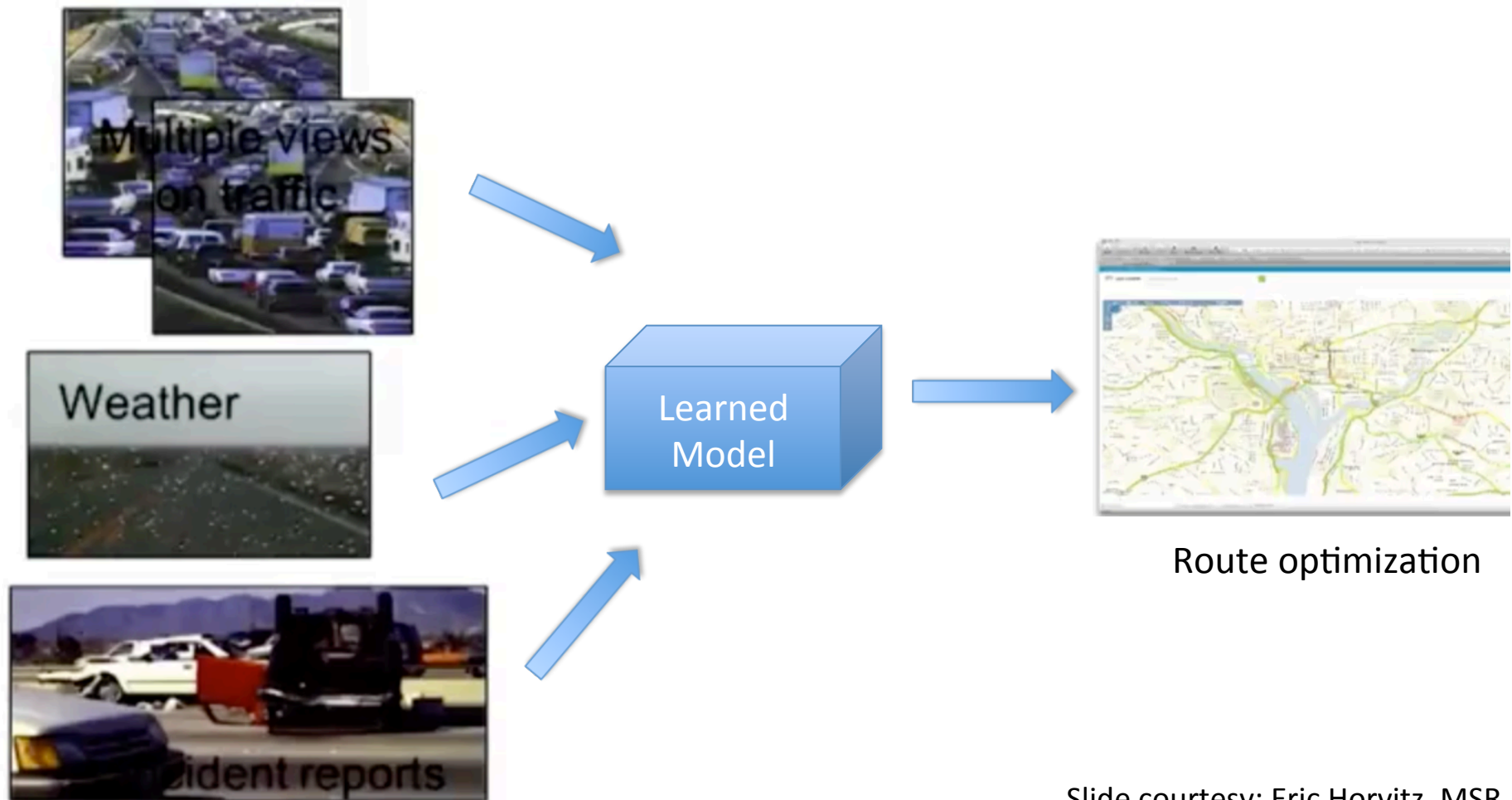
No graphical model



With graphical model

# Multi-sensor integration: Traffic

- Learn from historical data to make predictions



# Stock market

Google Inc (NASDAQ:GOOG)

Add to portfolio

More results

**744.00** +41.13 (5.85%)

Real-time: 10:43AM EST  
 NASDAQ real-time data - Disclaimer  
 Currency in USD

Range	735.79 - 747.99	Div/yield	-
52 week	556.52 - 774.38	EPS	32.46
Open	735.99	Shares	328.59M
Vol / Avg.	2.68M/2.28M	Beta	1.08
Mkt cap	244.39B	Inst. own	69%
P/E	22.91		

+1 5k

Dow Jones	13,758.94	0.34%	
Nasdaq	3,151.72	0.27%	
Technology		0.33%	
GOOG	744.00	5.85%	



- A** [Google Inc. \(GOOG\) Is Up Sharply On Q4 Results](#)  
RTT News - 1 hour ago
  - B** [Stocks to Watch: Google, Coach, Annie's](#)  
Wall Street Journal - 1 hour ago
  - C** [Google Inc \(GOOG\) Reports Strong Earnings, Shares Rise](#)  
ValueWalk - 3 hours ago
  - D** [Google 4th-Quarter Profits Increase as Ad Pricing Improves](#)  
NASDAQ - 15 hours ago
  - E** [Facebook Inc \(FB\)'s Social Graph Is a Google Inc \(GOOG\) Plus Killer](#)  
Insider Monkey - 16 hours ago
- [All news for Google Inc »](#) [Subscribe](#)

Events [Add GOOG to my calendars](#)

Apr 15, 2013

# Going global: Local ambiguity

- Text recognition

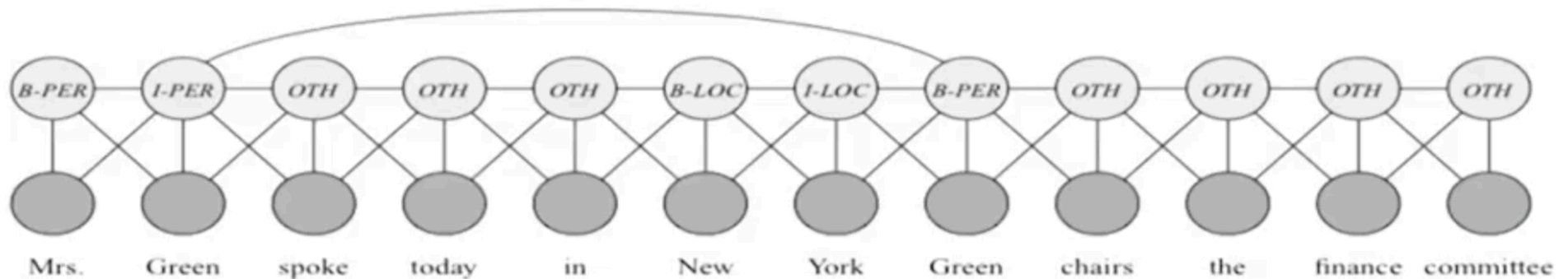
The image shows the words 'TAE' and 'CAT' in a stylized, blocky font. The letters are colored in two ways: green and red. In 'TAE', the 'T' and 'E' are green, while the 'A' is red. In 'CAT', the 'C' and 'T' are green, while the 'A' is red. This color scheme creates a visual ambiguity where the red 'A's in both words suggest a connection or a shared element, despite the words being different.

Smyth et al., 1994

# Going global: Local ambiguity

- Textual information extraction

e.g., Mrs. Green spoke today in New York. Green chairs the financial committee.



# Overview of the course

- Representation
  - How do we store  $P(Y_1, \dots, Y_n)$
  - Directed and undirected (model implications/assumptions)
- Inference
  - Answer questions with the model
  - Exact and approximate (marginal/most probable estimate)
- Learning
  - What model is right for data
  - Parameters and structure

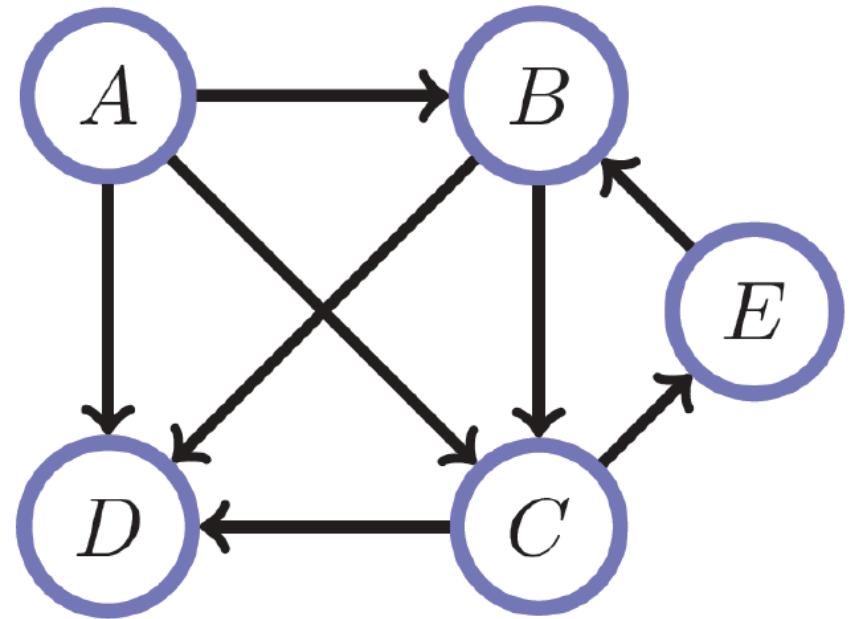
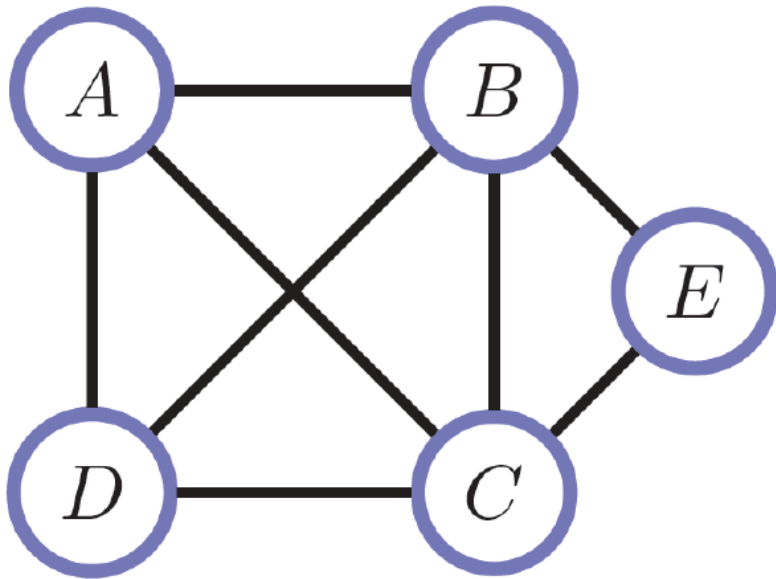
First, a recap of basics

# Graphs

- Concepts
  - Definition of G
  - Vertices/Nodes
  - Edges
  - Directed vs Undirected
  - Neighbours vs Parent/Child
  - Degree vs In/Out degree
  - Walk vs Path vs Cycle

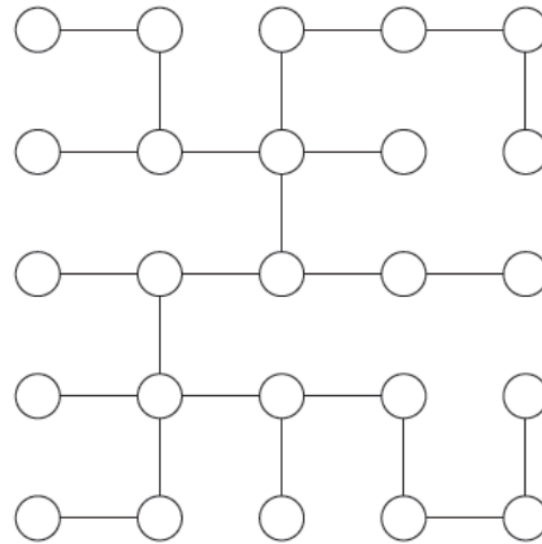
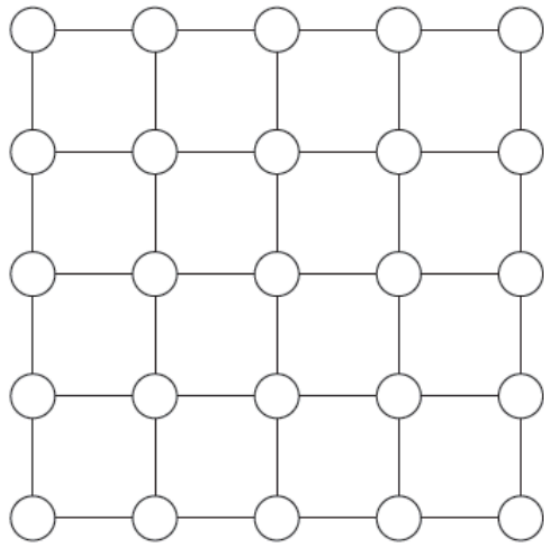


# Graphs

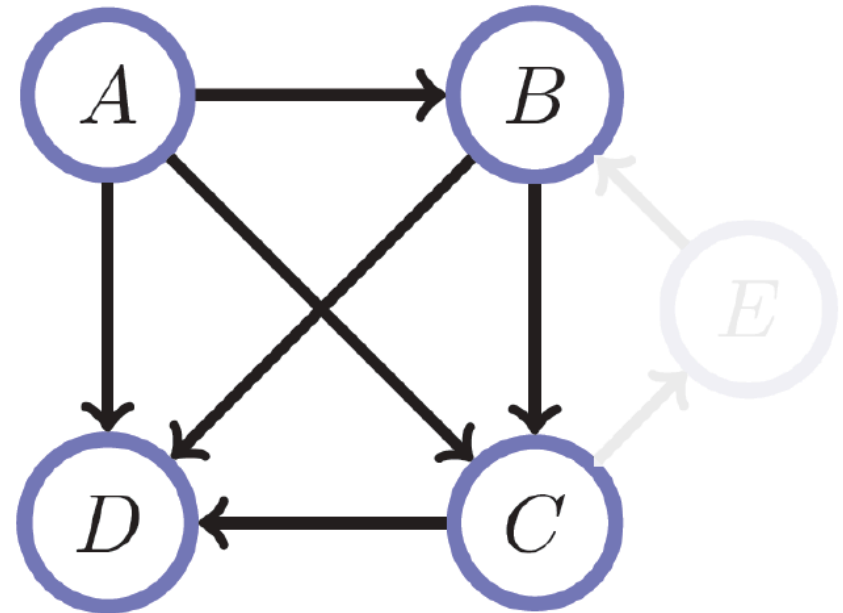
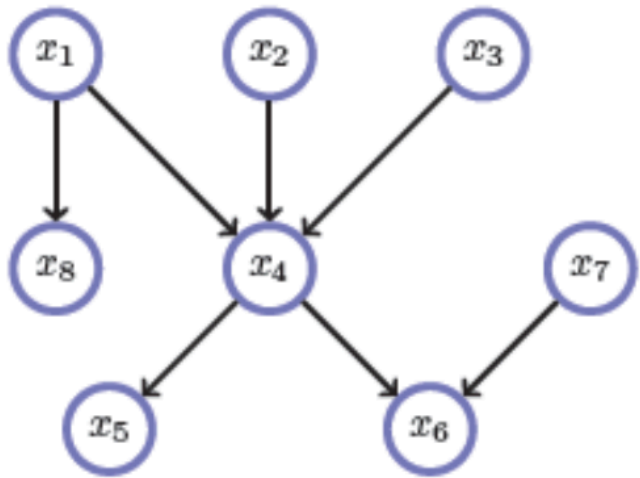


# Special graphs

- Trees: undirected graph, no cycles
- Spanning tree: Same set of vertices, but subset of edges, connected and no cycles



# Directed acyclic graphs (DAGs)



# Interpreting Probability

- What does  $P(A)$  mean?
- Frequentist view
  - Limit  $N \rightarrow \infty$ ,  $\#(A \text{ is true})/N$
  - i.e., limiting frequency of a repeating non-deterministic event
- Bayesian view
  - $P(A)$  is your belief about  $A$

# Joint distribution

- 3 variables
  - Intelligence (I)
  - Difficulty (D)
  - Grade (G)
- Independent parameters?

I	D	G	Prob.
$i^0$	$d^0$	$g^1$	0.126
$i^0$	$d^0$	$g^2$	0.168
$i^0$	$d^0$	$g^3$	0.126
$i^0$	$d^1$	$g^1$	0.009
$i^0$	$d^1$	$g^2$	0.045
$i^0$	$d^1$	$g^3$	0.126
$i^1$	$d^0$	$g^1$	0.252
$i^1$	$d^0$	$g^2$	0.0224
$i^1$	$d^0$	$g^3$	0.0056
$i^1$	$d^1$	$g^1$	0.06
$i^1$	$d^1$	$g^2$	0.036
$i^1$	$d^1$	$g^3$	0.024

# Conditioning

- Condition on  $g^1$

<b>I</b>	<b>D</b>	<b>G</b>	<b>Prob.</b>
$i^0$	$d^0$	$g^1$	0.126
$i^0$	$d^0$	$g^2$	0.168
$i^0$	$d^0$	$g^3$	0.126
$i^0$	$d^1$	$g^1$	0.009
$i^0$	$d^1$	$g^2$	0.045
$i^0$	$d^1$	$g^3$	0.126
$i^1$	$d^0$	$g^1$	0.252
$i^1$	$d^0$	$g^2$	0.0224
$i^1$	$d^0$	$g^3$	0.0056
$i^1$	$d^1$	$g^1$	0.06
$i^1$	$d^1$	$g^2$	0.036
$i^1$	$d^1$	$g^3$	0.024

# Conditioning

- $P(Y = y \mid X = x)$
- Informally,
  - What do you believe about  $Y=y$  when I tell you  $X=x$  ?
- $P(\text{France wins Euro 2020})$  ?
- What if I tell you:
  - France won the world cup 2018
  - Hasn't had catastrophic results since 😊

# Conditioning: Reduction

- Condition on  $g^1$

I	D	G	Prob.
$i^0$	$d^0$	$g^1$	0.126
$i^0$	$d^1$	$g^1$	0.009
$i^1$	$d^0$	$g^1$	0.252
$i^1$	$d^1$	$g^1$	0.06



# Conditioning: Renormalization

<b>I</b>	<b>D</b>	<b>G</b>	<b>Prob.</b>
$i^0$	$d^0$	$g^1$	0.126
$i^0$	$d^1$	$g^1$	0.009
$i^1$	$d^0$	$g^1$	0.252
$i^1$	$d^1$	$g^1$	0.06

$P(I, D, g^1)$

Unnormalized measure



<b>I</b>	<b>D</b>	<b>Prob.</b>
$i^0$	$d^0$	0.282
$i^0$	$d^1$	0.02
$i^1$	$d^0$	0.564
$i^1$	$d^1$	0.134

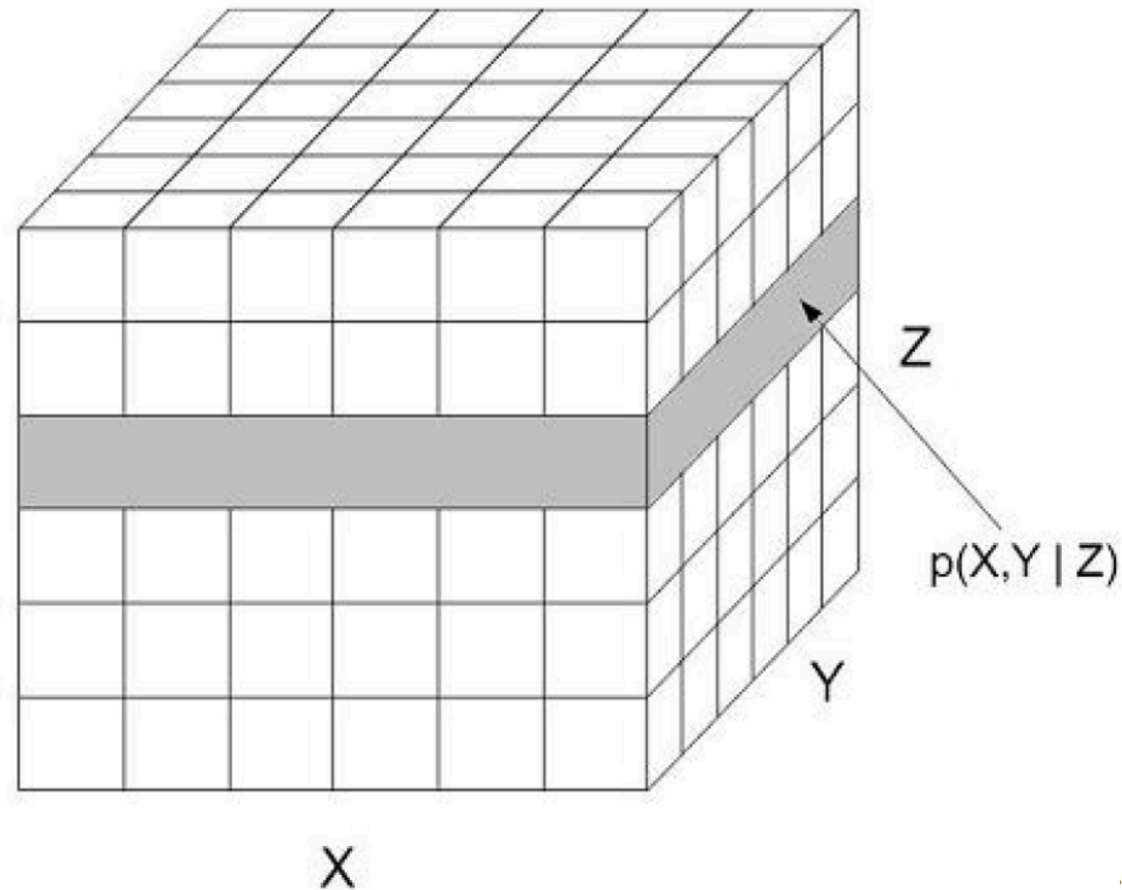
$P(I, D | g^1)$

# Conditional probability distribution

- Example  $P(G | I, D)$

	$g^1$	$g^2$	$g^3$
$i^0, d^0$	0.3	0.4	0.3
$i^0, d^1$	0.05	0.25	0.7
$i^1, d^0$	0.9	0.08	0.02
$i^1, d^1$	0.5	0.3	0.2

# Conditional probability distribution



$$p(x, y | Z = z) = \frac{p(x, y, z)}{p(z)}$$

# Marginalization

$P(I,D)$

<b>I</b>	<b>D</b>	<b>Prob.</b>
$i^0$	$d^0$	0.282
$i^0$	$d^1$	0.02
$i^1$	$d^0$	0.564
$i^1$	$d^1$	0.134

Marginalize I

<b>D</b>	<b>Prob.</b>
$d^0$	0.846
$d^1$	0.154

# Marginalization

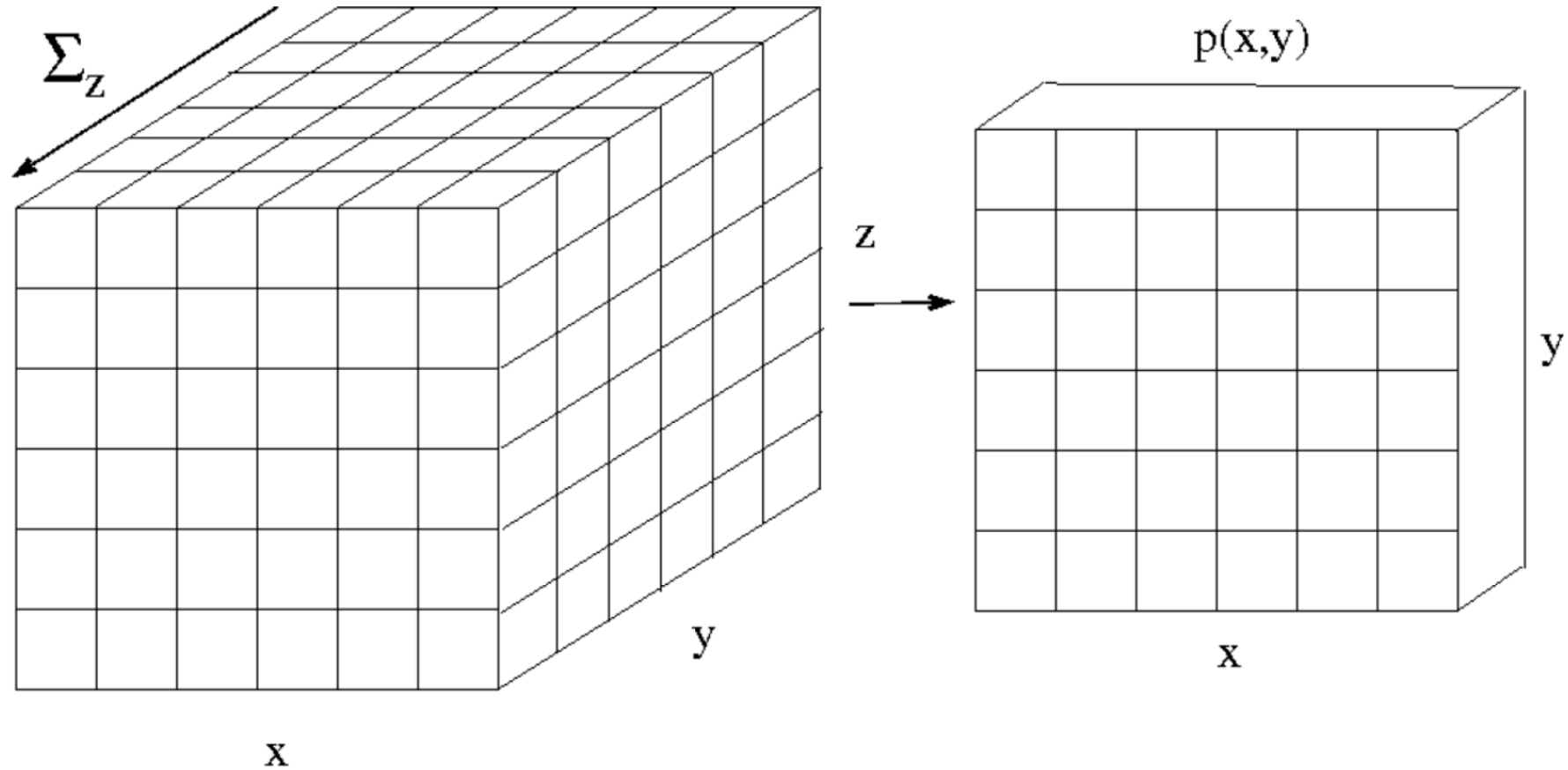
- Events

- $P(A) = P(A \text{ and } B) + P(A \text{ and not } B)$

- Random variables

- $P(X = x) = \sum_y P(X = x, Y = y)$

# Marginalization



$$p(x, y) = \sum_{z \in \mathcal{Z}} p(x, y, z)$$

$$p(x) = \sum_{y \in \mathcal{Y}} p(x, y)$$

Slide courtesy: Erik Sudderth

# Factors

- A factor  $\Phi(Y_1, \dots, Y_k)$

$$\Phi: \text{Val}(Y_1, \dots, Y_k) \rightarrow \mathbb{R}$$

- Scope =  $\{Y_1, \dots, Y_k\}$

# Factors

- Example:  $P(D, I, G)$

<b>I</b>	<b>D</b>	<b>G</b>	<b>Prob.</b>
$i^0$	$d^0$	$g^1$	0.126
$i^0$	$d^0$	$g^2$	0.168
$i^0$	$d^0$	$g^3$	0.126
$i^0$	$d^1$	$g^1$	0.009
$i^0$	$d^1$	$g^2$	0.045
$i^0$	$d^1$	$g^3$	0.126
$i^1$	$d^0$	$g^1$	0.252
$i^1$	$d^0$	$g^2$	0.0224
$i^1$	$d^0$	$g^3$	0.0056
$i^1$	$d^1$	$g^1$	0.06
$i^1$	$d^1$	$g^2$	0.036
$i^1$	$d^1$	$g^3$	0.024



# Factors

- Example:  $P(D, I, g^1)$

<b>I</b>	<b>D</b>	<b>G</b>	<b>Prob.</b>
$i^0$	$d^0$	$g^1$	0.126
$i^0$	$d^1$	$g^1$	0.009
$i^1$	$d^0$	$g^1$	0.252
$i^1$	$d^1$	$g^1$	0.06

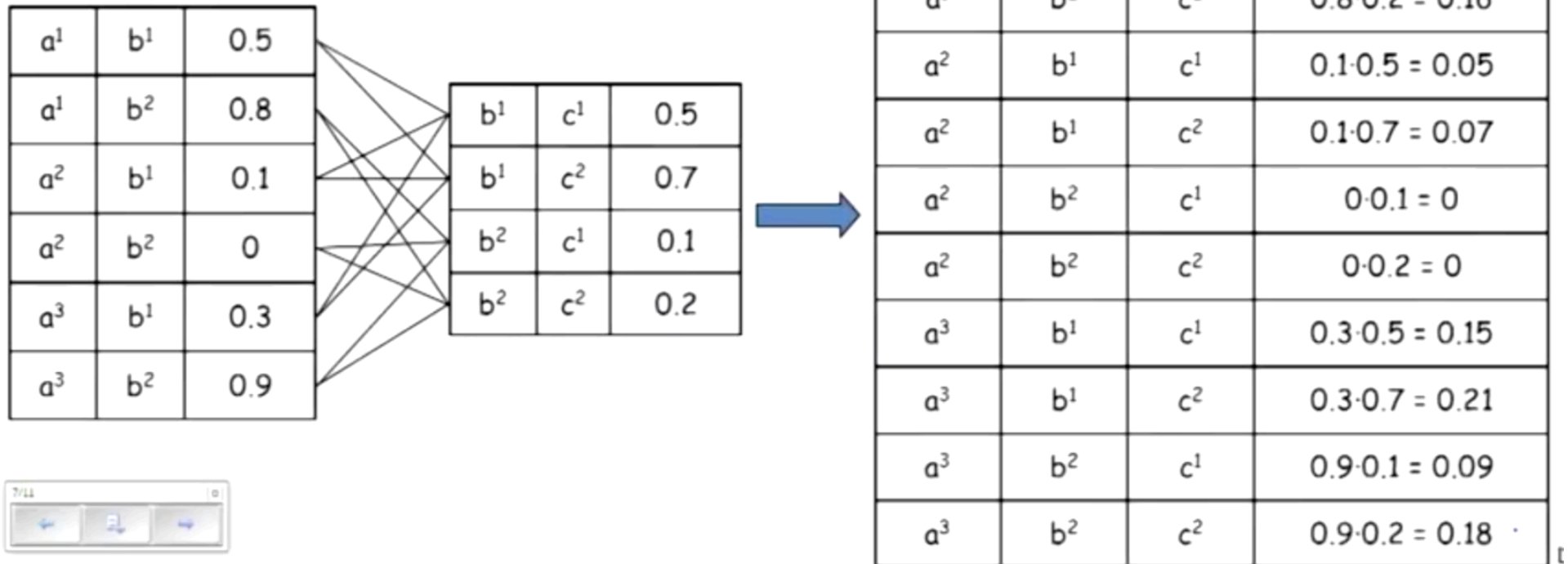
What is the scope here?

# General factors

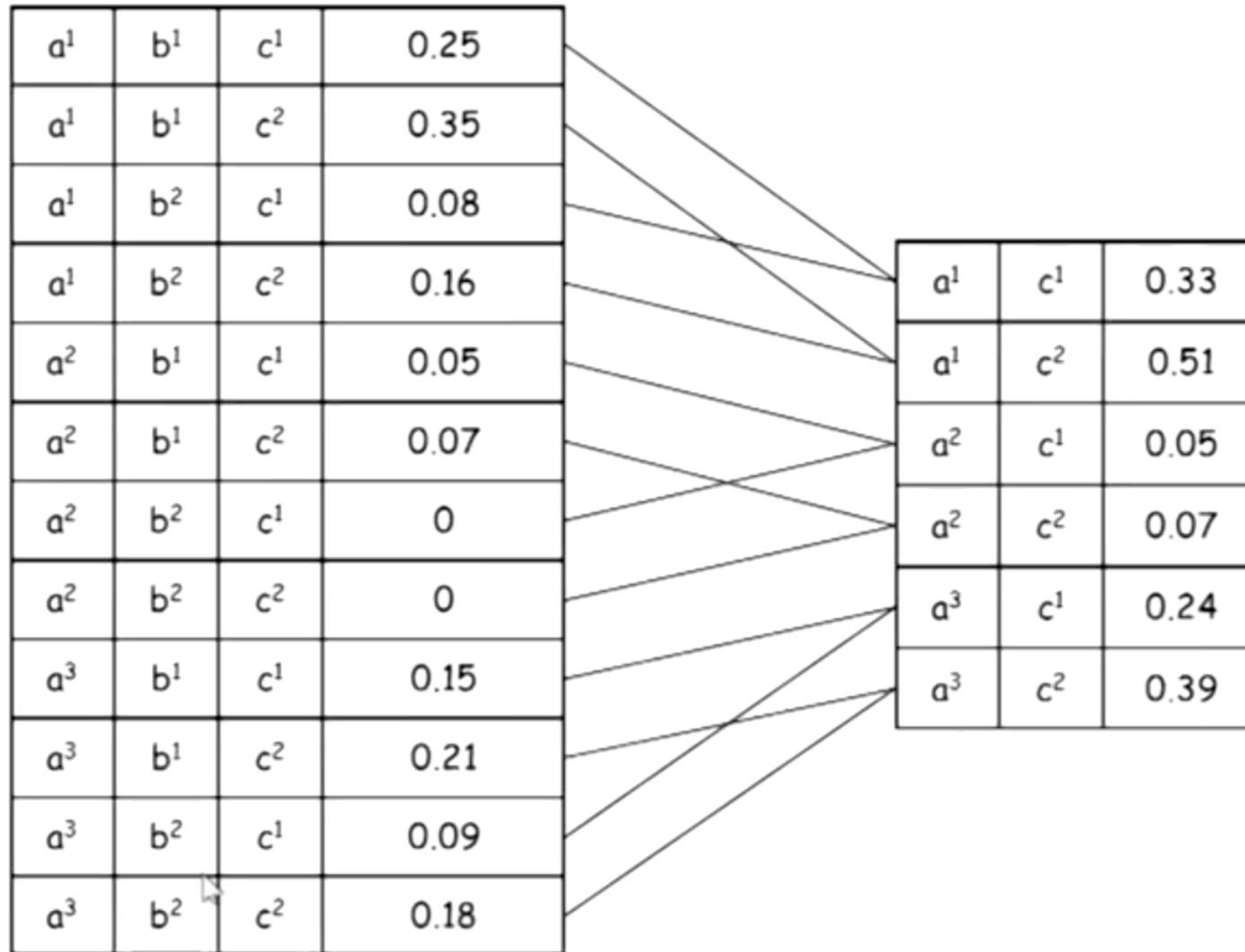
- Not necessarily for probabilities

<b>A</b>	<b>B</b>	$\phi$
$a^0$	$b^0$	30
$a^0$	$b^1$	5
$a^1$	$b^0$	1
$a^1$	$b^1$	10

# Factor product



# Factor marginalization



# Factor reduction

$a^1$	$b^1$	$c^1$	0.25
$a^1$	$b^1$	$c^2$	0.35
$a^1$	$b^2$	$c^1$	0.08
$a^1$	$b^2$	$c^2$	0.16
$a^2$	$b^1$	$c^1$	0.05
$a^2$	$b^1$	$c^2$	0.07
$a^2$	$b^2$	$c^1$	0
$a^2$	$b^2$	$c^2$	0
$a^3$	$b^1$	$c^1$	0.15
$a^3$	$b^1$	$c^2$	0.21
$a^3$	$b^2$	$c^1$	0.09
$a^3$	$b^2$	$c^2$	0.18

$a^1$	$b^1$	$c^1$	0.25
$a^1$	$b^2$	$c^1$	0.08
$a^2$	$b^1$	$c^1$	0.05
$a^2$	$b^2$	$c^1$	0
$a^3$	$b^1$	$c^1$	0.15
$a^3$	$b^2$	$c^1$	0.09

# Why factors ?

- Building blocks for defining distributions in high-dimensional spaces
- Set of basic operations for manipulating these distributions

# Independent random variables

$P(x,y)$


=


--	--	--	--	--

$X \perp Y$



$$p(x, y) = p(x)p(y)$$

for all  $x \in \mathcal{X}, y \in \mathcal{Y}$

# Marginal independence

- **Sets** of variables  $\mathbf{X}$ ,  $\mathbf{Y}$
- $\mathbf{X}$  is independent of  $\mathbf{Y}$ 
  - Shorthand:  $P \models (\mathbf{X} \perp \mathbf{Y})$
- **Proposition:**  $P$  satisfies  $(\mathbf{X} \perp \mathbf{Y})$  if and only if
  - $P(\mathbf{X}=\mathbf{x}, \mathbf{Y}=\mathbf{y}) = P(\mathbf{X}=\mathbf{x}) P(\mathbf{Y}=\mathbf{y}), \quad \forall \mathbf{x} \in \text{Val}(\mathbf{X}), \mathbf{y} \in \text{Val}(\mathbf{Y})$



# Conditional independence

- **Sets** of variables  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$
- $\mathbf{X}$  is independent of  $\mathbf{Y}$  given  $\mathbf{Z}$  if
  - Shorthand:  $P \vdash (\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$
  - For  $P \vdash (\mathbf{X} \perp \mathbf{Y} \mid \emptyset)$ , write  $P \vdash (\mathbf{X} \perp \mathbf{Y})$
- **Proposition:**  $P$  satisfies  $(\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z})$  if and only if
  - $P(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Z}) P(\mathbf{Y} \mid \mathbf{Z})$ ,  $\forall \mathbf{x} \in \text{Val}(\mathbf{X}), \mathbf{y} \in \text{Val}(\mathbf{Y}), \mathbf{z} \in \text{Val}(\mathbf{Z})$

# Bayes Rule

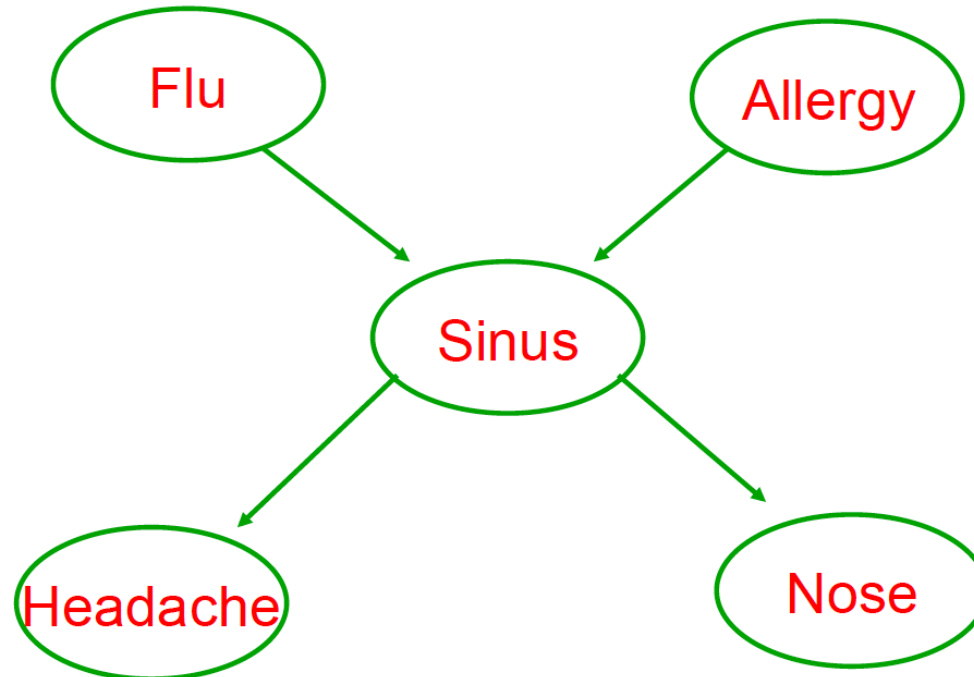
- Simple yet profound
- Concepts
  - Likelihood
    - How much does a certain hypothesis explain the data?
  - Prior
    - What do you believe before seeing any data?
  - Posterior
    - What do we believe after seeing the data?

# Bayesian Networks

- DAGs
  - nodes represent variables in the Bayesian sense
  - edges represent conditional dependencies
- Example
  - Suppose that we know the following:
    - The flu causes sinus inflammation
    - Allergies cause sinus inflammation
    - Sinus inflammation causes a runny nose
    - Sinus inflammation causes headaches
  - How are these connected ?

# Bayesian Networks

- Example



# Bayesian Networks

- A general Bayes net
  - Set of random variables
  - DAG: encodes independence assumptions
  - Conditional probability trees
  - Joint distribution

$$P(Y_1, \dots, Y_n) = \prod_{i=1}^n P(Y_i \mid \text{Pa}_{Y_i})$$

# Bayesian Networks

- A general Bayes net
  - How many parameters ?
    - Discrete variables  $Y_1, \dots, Y_n$
    - Graph: Defines parents of  $Y_i$ , i.e.,  $(Pa_{Y_i})$
    - CPTs:  $P(Y_i | Pa_{Y_i})$

# Markov nets

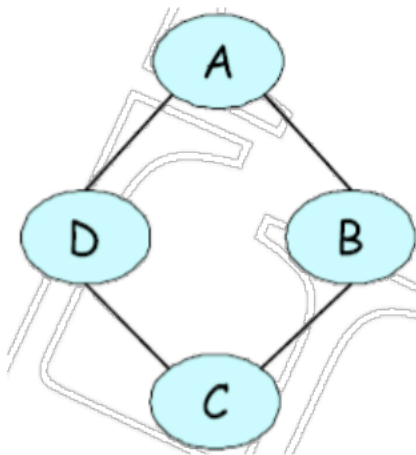
- Set of random variables
- Undirected graph
  - Encodes independence assumptions
- Factors

Comparison to Bayesian Nets ?

# Pairwise MRFs

- Composed of pairwise factors
  - A function of two variables
  - Can also have unary terms

- Example



$\phi_1[A, B]$			$\phi_2[B, C]$			$\phi_3[C, D]$			$\phi_4[D, A]$		
$a^0$	$b^0$	30	$b^0$	$c^0$	100	$c^0$	$d^0$	1	$d^0$	$a^0$	100
$a^0$	$b^1$	5	$b^0$	$c^1$	1	$c^0$	$d^1$	100	$d^0$	$a^1$	1
$a^1$	$b^0$	1	$b^1$	$c^0$	1	$c^1$	$d^0$	100	$d^1$	$a^0$	1
$a^1$	$b^1$	10	$b^1$	$c^1$	100	$c^1$	$d^1$	1	$d^1$	$a^1$	100



# Markov Nets: Computing probabilities

- Can only compute ratio of probabilities directly

$\phi_1[A, B]$			$\phi_2[B, C]$			$\phi_3[C, D]$			$\phi_4[D, A]$		
$a^0$	$b^0$	30	$b^0$	$c^0$	100	$c^0$	$d^0$	1	$d^0$	$a^0$	100
$a^0$	$b^1$	5	$b^0$	$c^1$	1	$c^0$	$d^1$	100	$d^0$	$a^1$	1
$a^1$	$b^0$	1	$b^1$	$c^0$	1	$c^1$	$d^0$	100	$d^1$	$a^0$	1
$a^1$	$b^1$	10	$b^1$	$c^1$	100	$c^1$	$d^1$	1	$d^1$	$a^1$	100

- Need to normalize with a **partition function**
  - Hard ! (sum over all possible assignments)
- In Bayesian Nets, can do by multiplying CPTs

# Markov nets $\leftrightarrow$ Factorization

- Given an undirected graph  $H$  over variables  $Y = \{Y_1, \dots, Y_n\}$
- A distribution  $P$  factorizes over  $H$  if there exist
  - Subsets of variables  $S^i \subseteq Y$  s.t.  $S^i$  are fully-connected in  $H$
  - Non-negative potentials (factors)  $\Phi_1(S^1), \dots, \Phi_m(S^m)$ : clique potentials
  - Such that

$$P(Y_1, \dots, Y_n) = \frac{1}{Z} \prod_{i=1}^m \Phi_i(S^i)$$

# Conditional Markov Random Fields

- Also known as: Markov networks, undirected graphical models, MRFs
- Note: Not making a distinction between CRFs and MRFs
- $\mathbf{X} \in \mathcal{X}$  : observed random variables
- $\mathbf{Y} = (Y_1, \dots, Y_n) \in \mathcal{Y}$ : output random variables
- $\mathbf{Y}_c$  are subset of variables for clique  $c \subseteq \{1, \dots, n\}$
- Define a factored probability distribution

$$P(\mathbf{Y} | \mathbf{X}) = \frac{1}{Z(\mathbf{X})} \prod_c \psi_c(\mathbf{Y}_c; \mathbf{X})$$

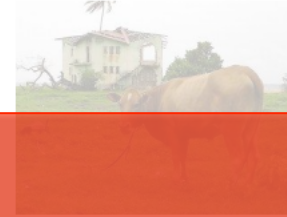
Partition function =  $\sum_{\mathbf{Y} \in \mathcal{Y}} \prod_c \psi_c(\mathbf{Y}_c; \mathbf{X})$  **Exponential number of configurations !**

# MRFs / CRFs

- Several applications, e.g., computer vision



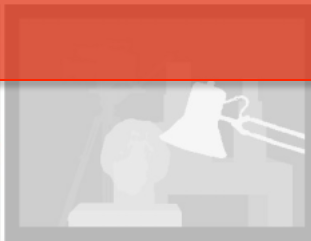
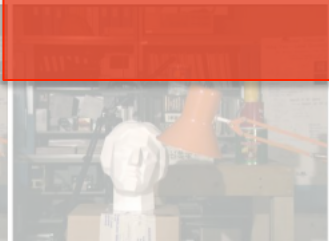
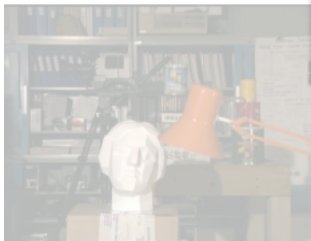
Interactive figure-ground segmentation [Boykov and Jolly, 2001; Boykov and Funka-Lea, 2003]



Surface context [Hoiem et al., 2005]

Semantic labeling [He et al., 2004; Shotton et al., 2006; Gould et al., 2005]

## Low-level vision problems



Stereo matching [Kolmogorov and Zabih, 2001; Scharstein and Szeliski, 2002]

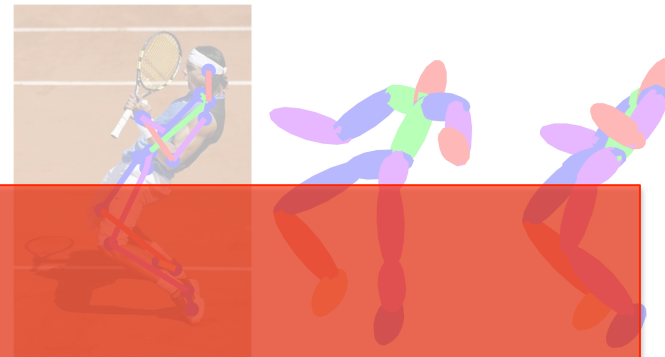
Image denoising [Felzenszwalb and Huttenlocher 2004]

# MRFs / CRFs

- Several applications, e.g., computer vision

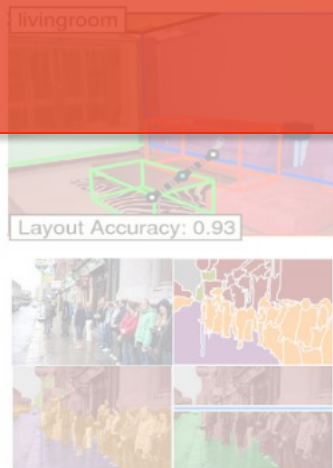
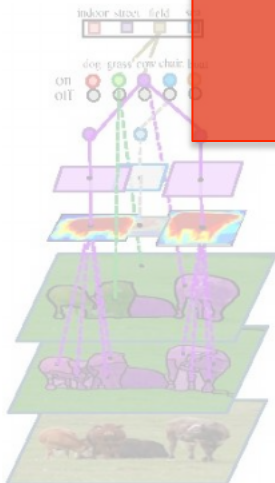


Object detection [Felzenszfeld et al., 2008]



Human pose estimation [Lipton and Black, 2015; Ramakrishna et al., 2012]

## High-level vision problems

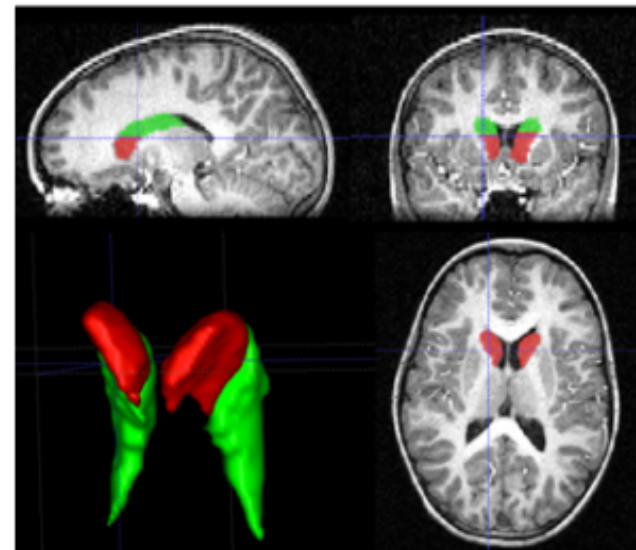
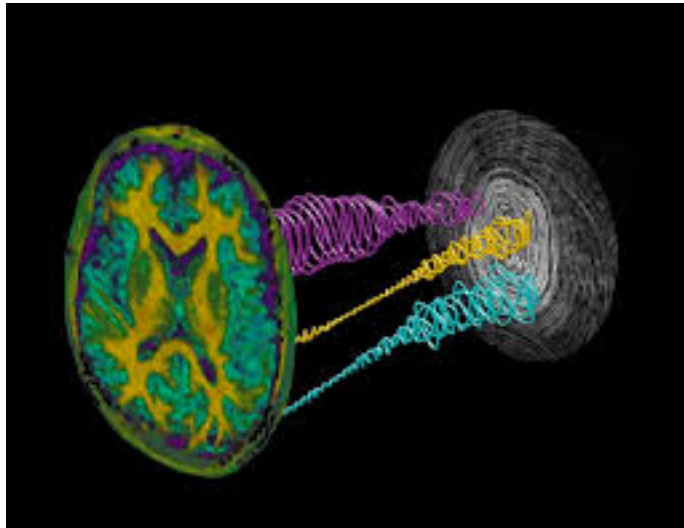


Scene understanding

[Fouhey et al., 2014; Ladicky et al., 2010; Xiao et al., 2013; Yao et al., 2012]

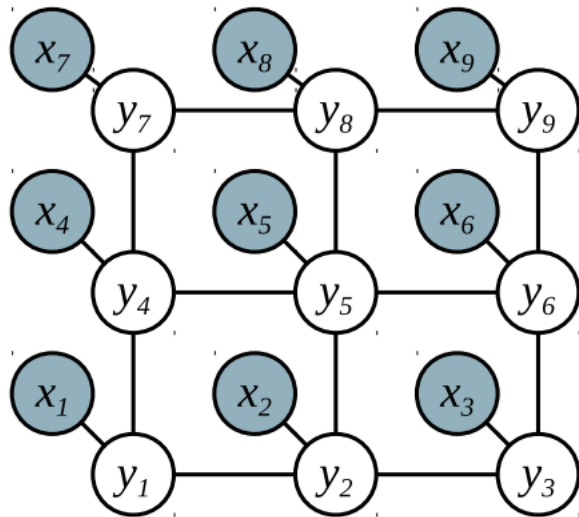
# MRFs / CRFs

- Several applications, e.g., medical imaging

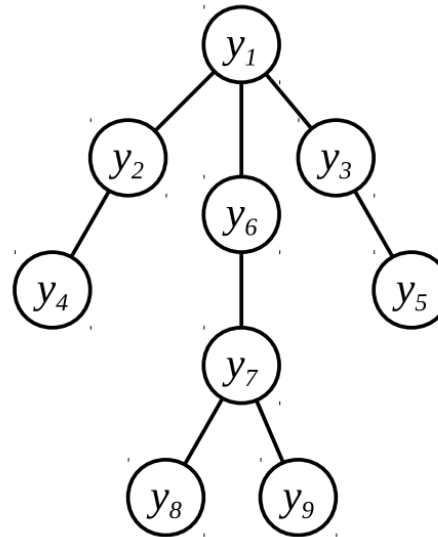


# MRFs / CRFs

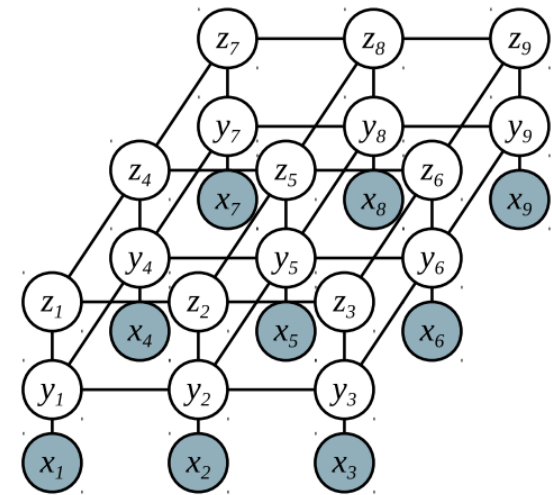
- Inherent in all these problems are graphical models



Pixel labeling



Object detection  
Pose estimation



Scene understanding

# Maximum a posteriori (MAP) inference

$$\begin{aligned} \mathbf{y}^* &= \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{y} \mid \mathbf{x}) \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \frac{1}{Z(\mathbf{x})} \prod_c \psi_c(\mathbf{Y}_c; \mathbf{X}) \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \log \left( \frac{1}{Z(\mathbf{x})} \prod_c \psi_c(\mathbf{Y}_c; \mathbf{X}) \right) \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \sum_c \log \psi_c(\mathbf{Y}_c; \mathbf{X}) - \log Z(\mathbf{X}) \\ &= \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \sum_c \log \psi_c(\mathbf{Y}_c; \mathbf{X}) \rightarrow -E(\mathbf{Y}; \mathbf{X}) \end{aligned}$$



# Maximum a posteriori (MAP) inference

$$\begin{aligned}\mathbf{y}^* &= \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{y} \mid \mathbf{x}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \sum_c \log \Psi_c(\mathbf{Y}_c; \mathbf{X}) \\ &= \operatorname{argmin}_{\mathbf{y} \in \mathcal{Y}} E(\mathbf{y}; \mathbf{x})\end{aligned}$$

MAP inference  $\Leftrightarrow$  Energy minimization

The energy function is  $E(\mathbf{Y}; \mathbf{X}) = \sum_c \psi_c(\mathbf{Y}_c; \mathbf{X})$

where  $\psi_c(\cdot) = -\log \Psi_c(\cdot)$

 Clique potential

# Clique potentials

- Defines a mapping from an assignment of random variables to a real number

$$\psi_c : \mathcal{Y}_c \times \mathcal{X} \rightarrow \mathbb{R}$$

- Encodes a preference for assignments to the random variables (lower is better)

- Parameterized as  $\psi_c(\mathbf{y}_c; \mathbf{x}) = \mathbf{w}_c^T \phi_c(\mathbf{y}_c; \mathbf{x})$

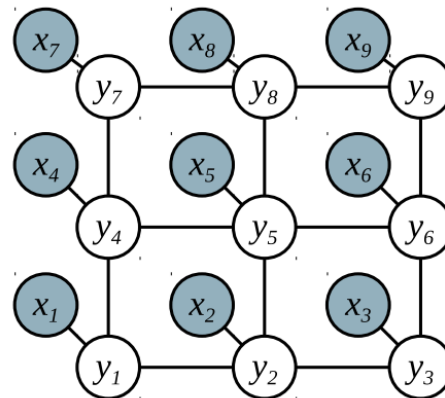


Parameters

# Clique potentials

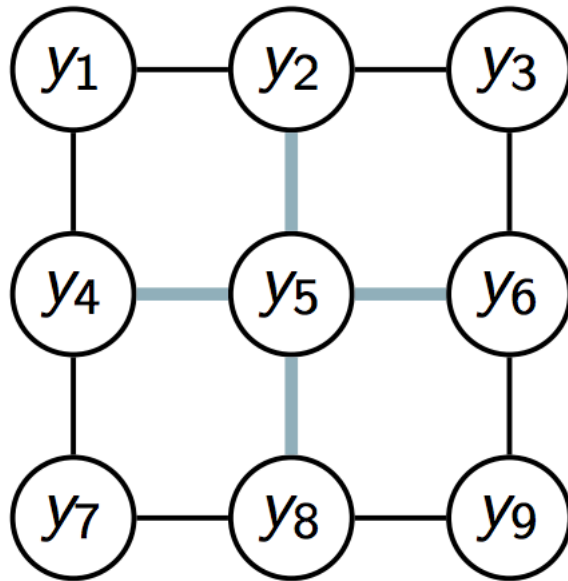
- Arity

$$\begin{aligned} E(\mathbf{y}; \mathbf{x}) &= \sum_c \psi_c(\mathbf{y}_c; \mathbf{x}) \\ &= \underbrace{\sum_{i \in \mathcal{V}} \psi_i^U(y_i; \mathbf{x})}_{\text{unary}} + \underbrace{\sum_{ij \in \mathcal{E}} \psi_{ij}^P(y_i, y_j; \mathbf{x})}_{\text{pairwise}} + \underbrace{\sum_{c \in \mathcal{C}} \psi_c^H(\mathbf{y}_c; \mathbf{x})}_{\text{higher-order}}. \end{aligned}$$

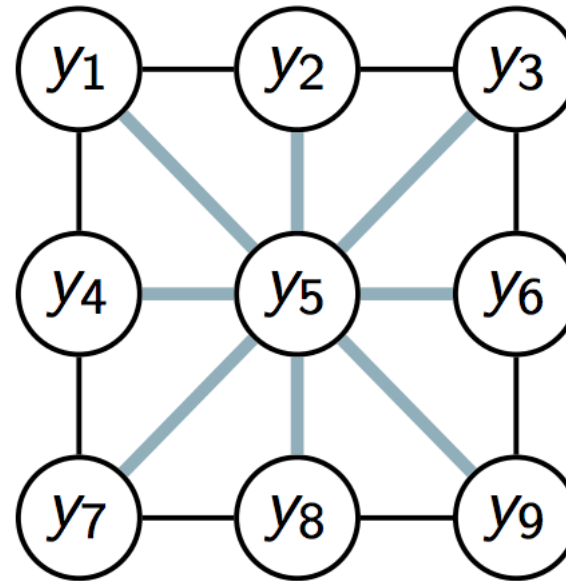


# Clique potentials

- Arity

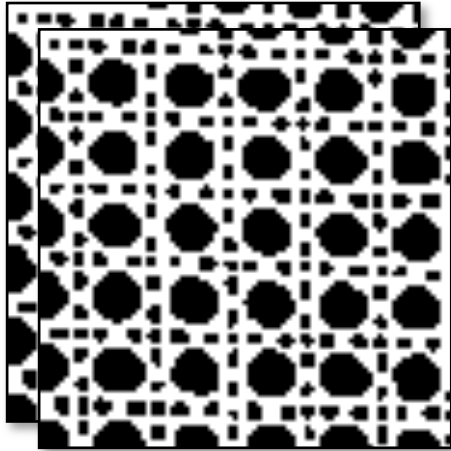


4-connected,  $\mathcal{N}_4$

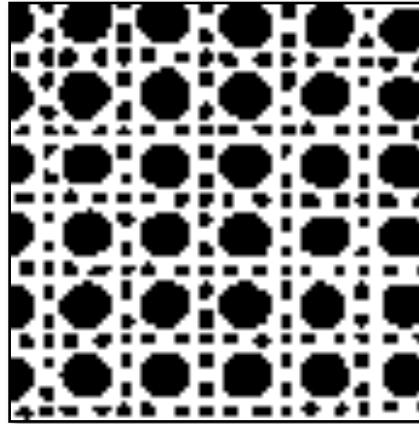


8-connected,  $\mathcal{N}_8$

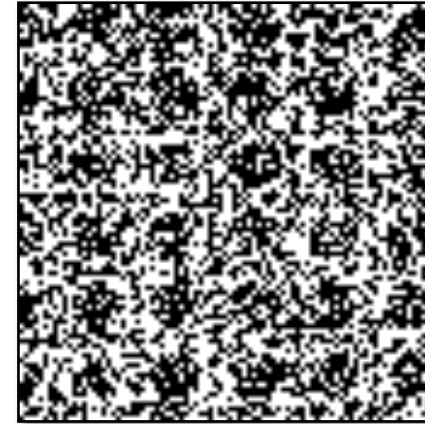
# Reason 1: Texture modelling



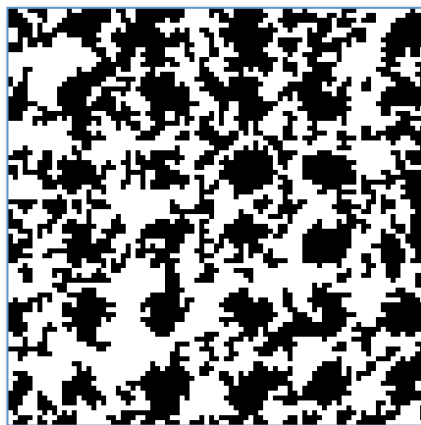
Training images



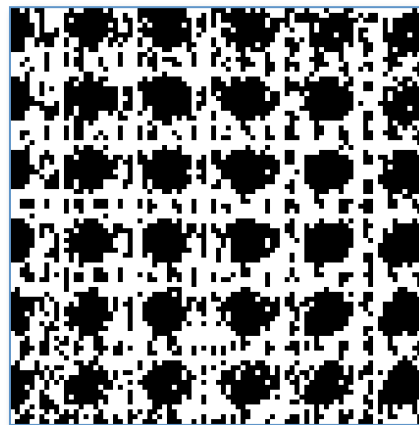
Test image



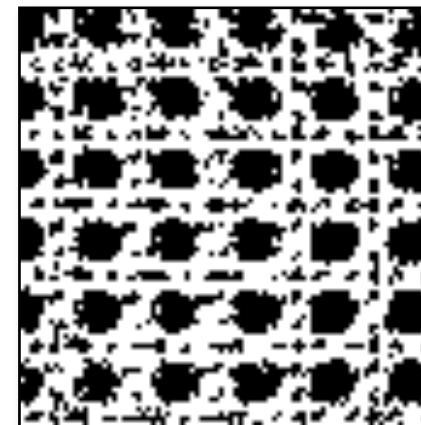
Test image (60% Noise)



Result MRF  
4-connected  
(neighbours)

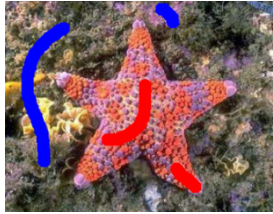


Result MRF  
4-connected

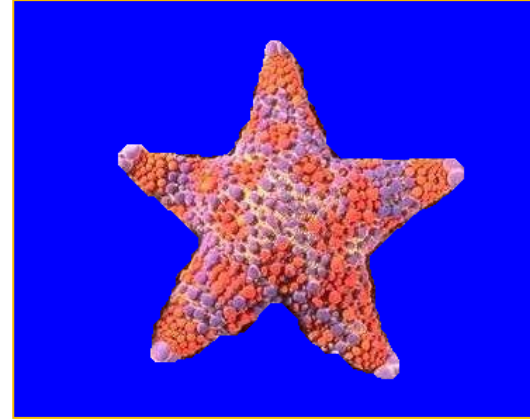


Result MRF  
9-connected  
(7 attractive; 2 repulsive)

# Reason2: Discretization artefacts



4-connected  
Euclidean



8-connected  
Euclidean

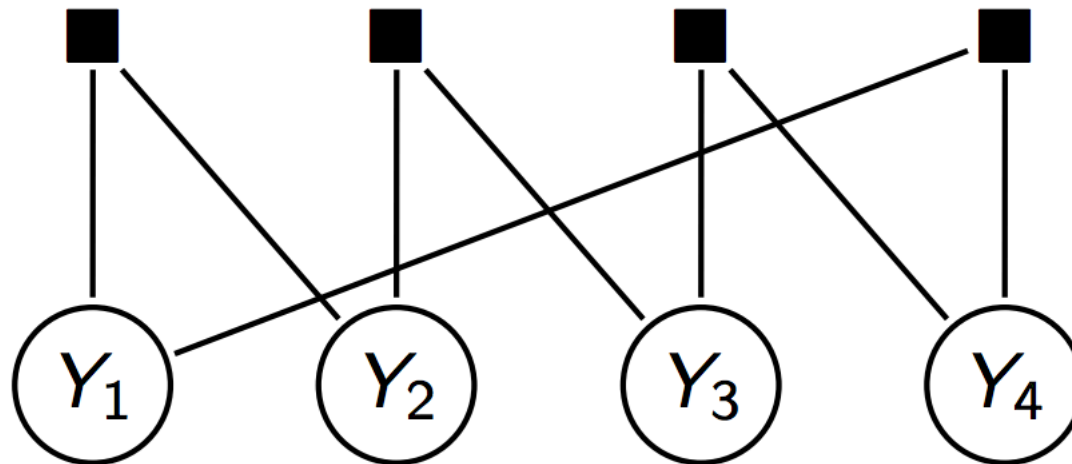
higher-connectivity can model  
true Euclidean length

[Boykov et al. '03; '05]

# Graphical representation

- Example

$$E(\mathbf{y}) = \psi(y_1, y_2) + \psi(y_2, y_3) + \psi(y_3, y_4) + \psi(y_4, y_1)$$

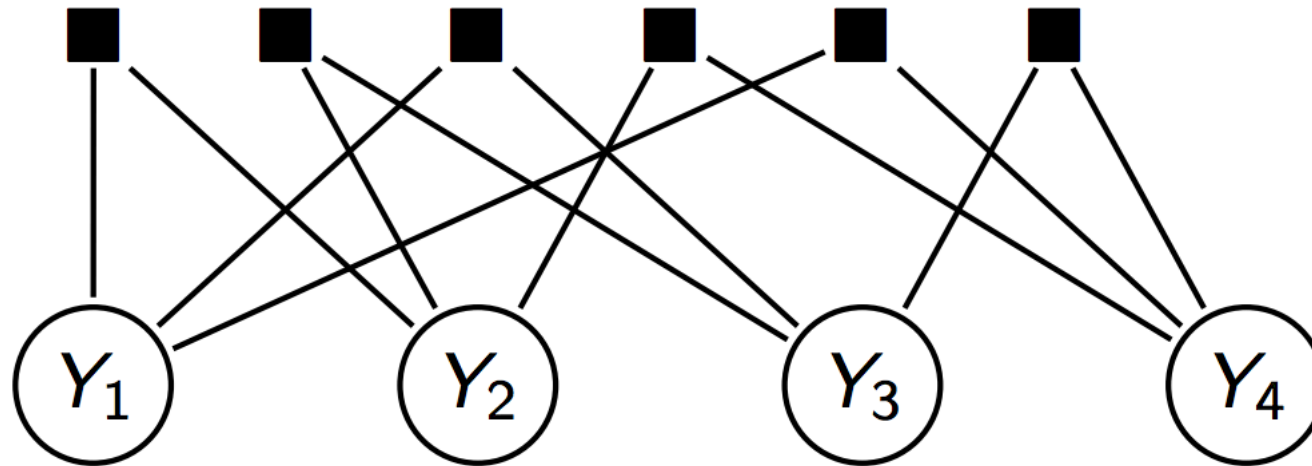


factor graph

# Graphical representation

- Example

$$E(\mathbf{y}) = \sum_{i,j} \psi(y_i, y_j)$$



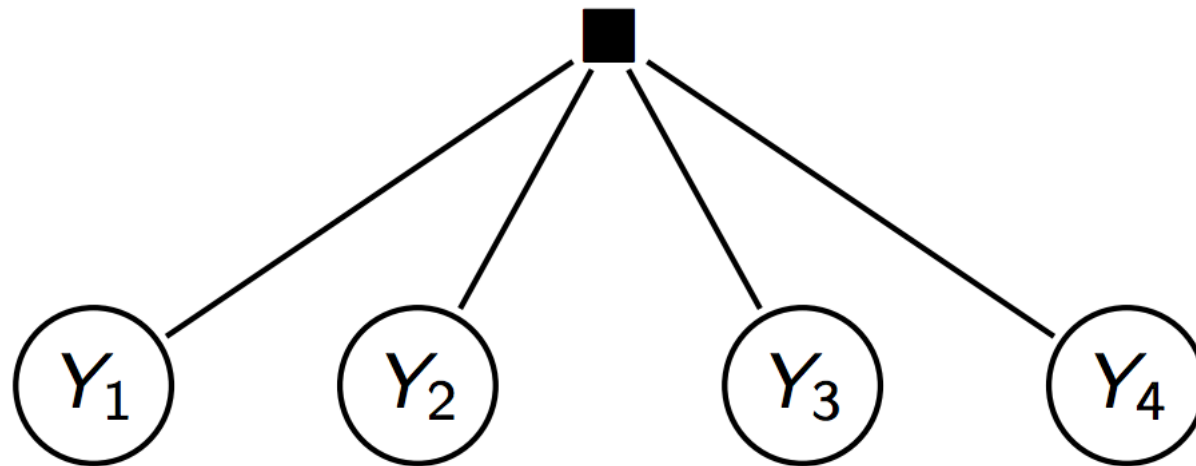
factor graph



# Graphical representation

- Example

$$E(\mathbf{y}) = \psi(y_1, y_2, y_3, y_4)$$



factor graph

# A Computer Vision Application

## Binary Image Segmentation



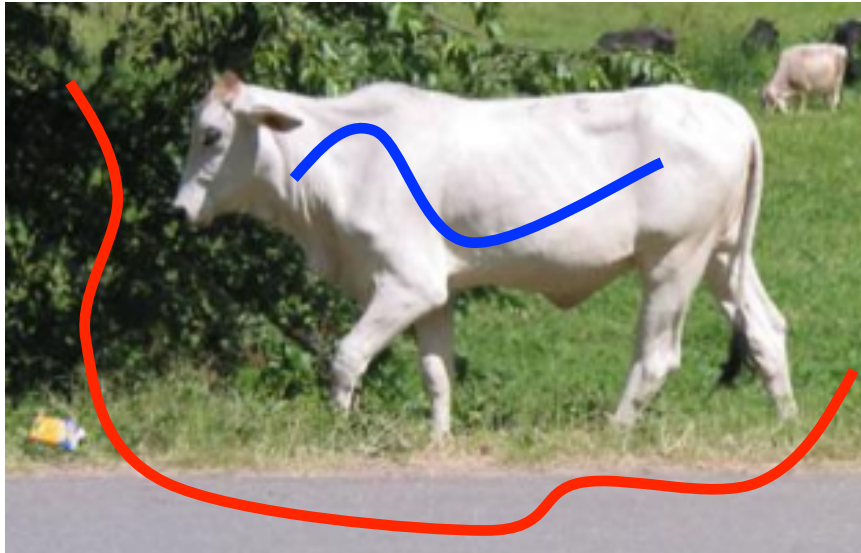
**How ?**

Cost function    Models *our* knowledge about natural images

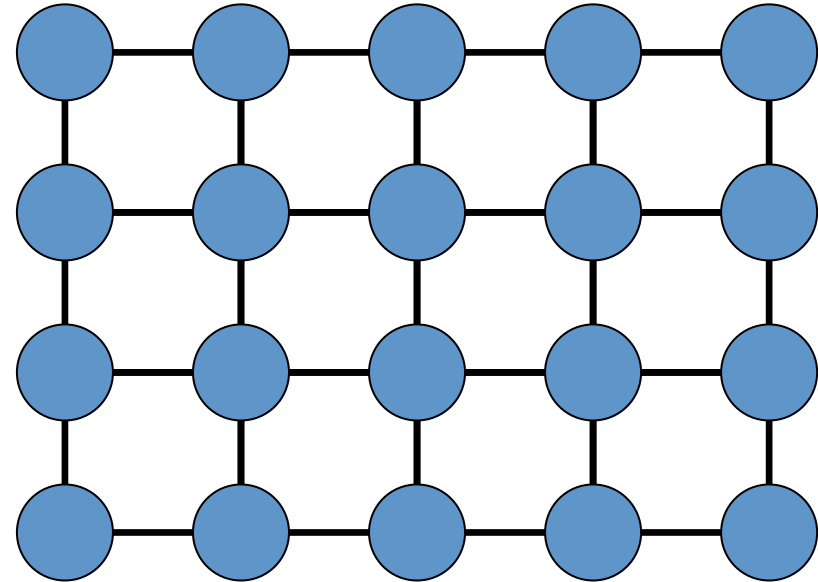
Optimize cost function to obtain the segmentation

# A Computer Vision Application

## Binary Image Segmentation



Object - white, Background - green/grey



Graph  $G = (V, E)$

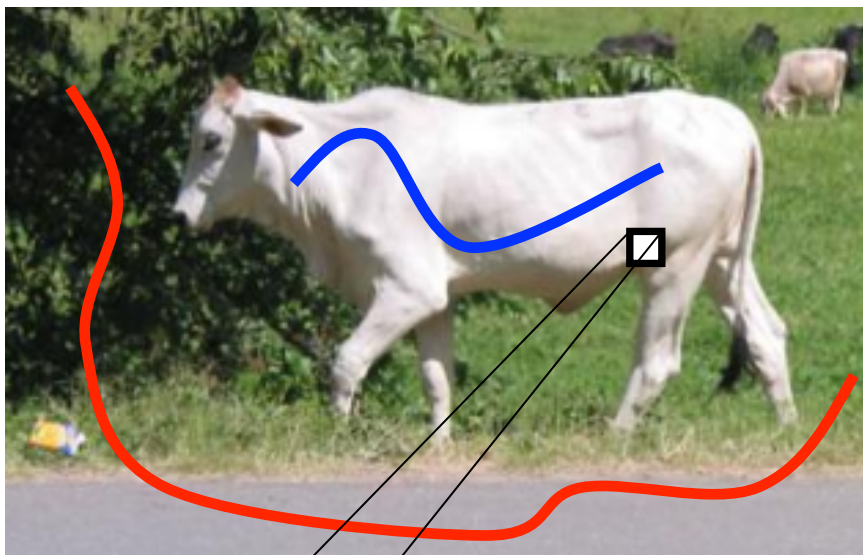
Each vertex corresponds to a pixel

Edges define a 4-neighbourhood *grid* graph

Assign a label to each vertex from  $L = \{\text{obj}, \text{bkg}\}$

# A Computer Vision Application

## Binary Image Segmentation

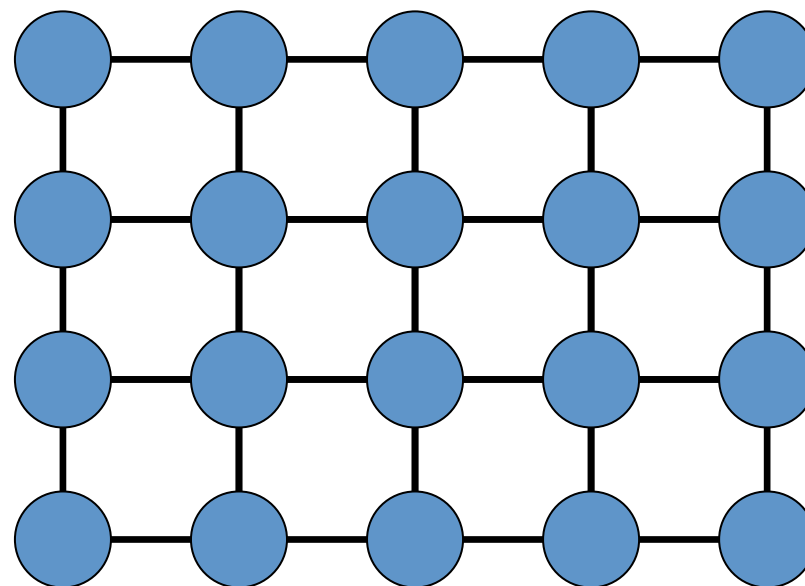


Object - white, Background - green/grey

Cost of a labelling  $f : V \rightarrow L$



Cost of label 'obj' low Cost of label 'bkg' high

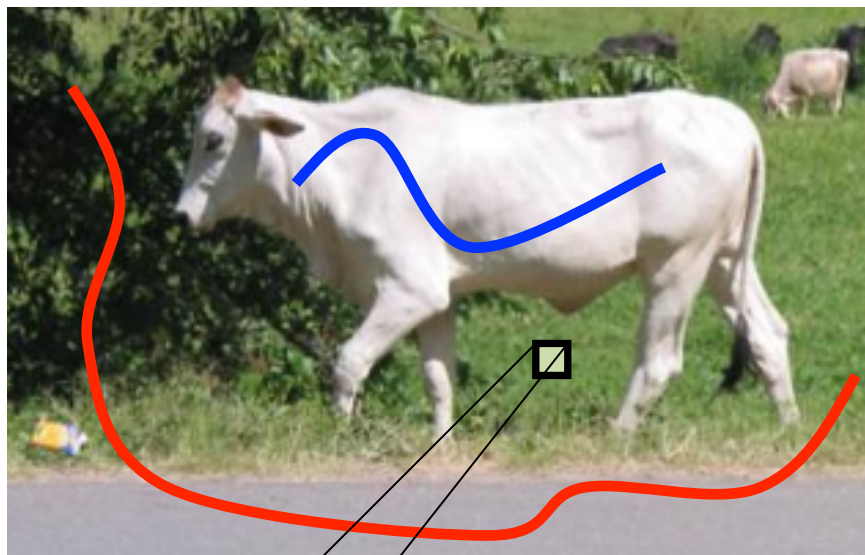


Graph  $G = (V, E)$

Per Vertex Cost

# A Computer Vision Application

## Binary Image Segmentation

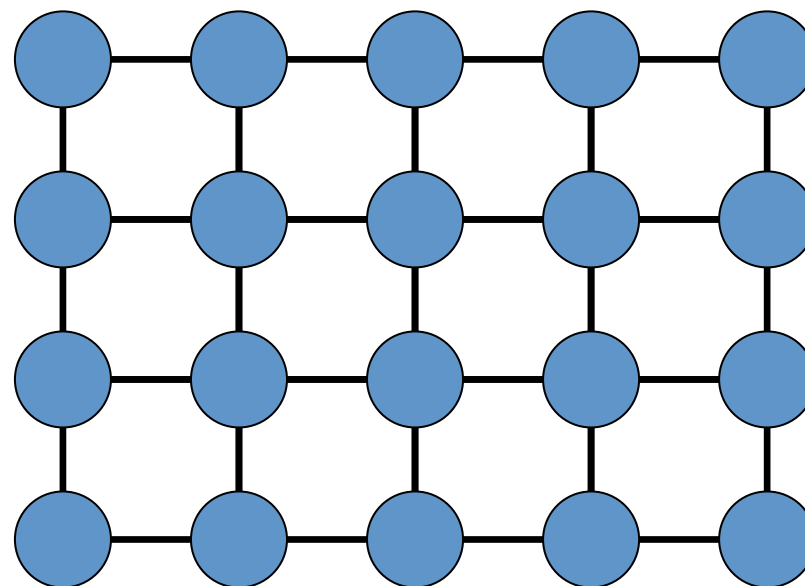


Object - white, Background - green/grey

Cost of a labelling  $f : V \rightarrow L$



Cost of label 'obj' high Cost of label 'bkg' low



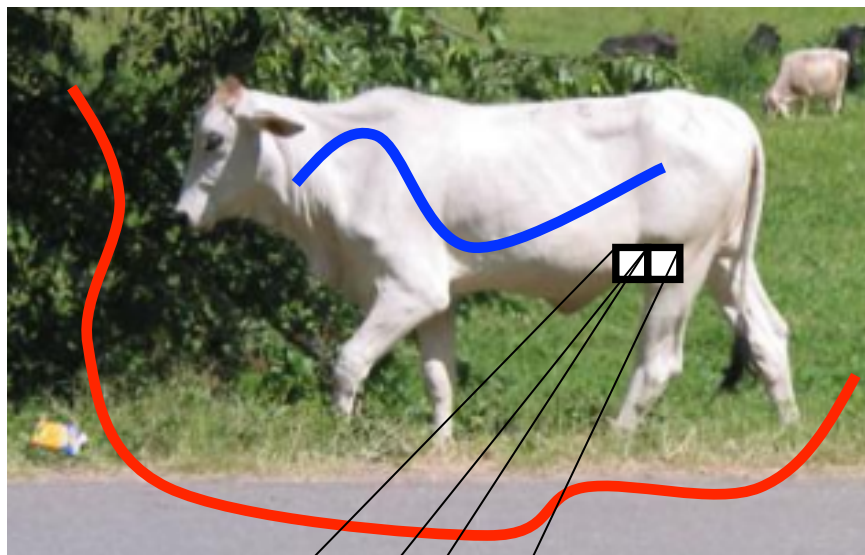
Graph  $G = (V, E)$

Per Vertex Cost

UNARY COST

# A Computer Vision Application

## Binary Image Segmentation



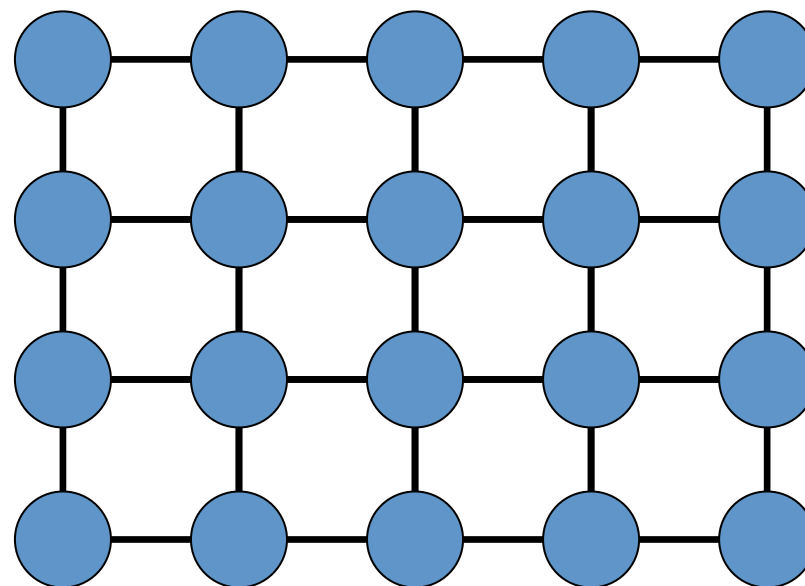
Object - white, Background - green/grey

Cost of a labelling  $f : V \rightarrow L$



Cost of same label low

Cost of different labels high

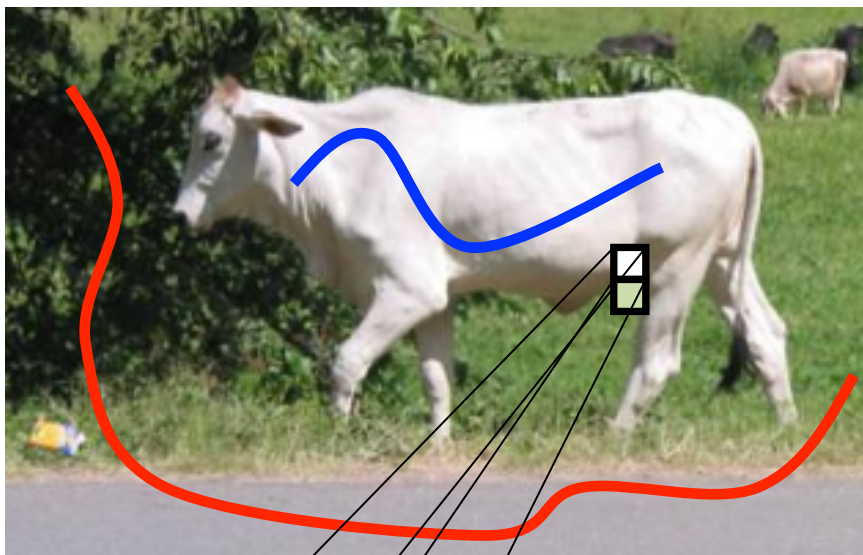


Graph  $G = (V, E)$

Per Edge Cost

# A Computer Vision Application

## Binary Image Segmentation



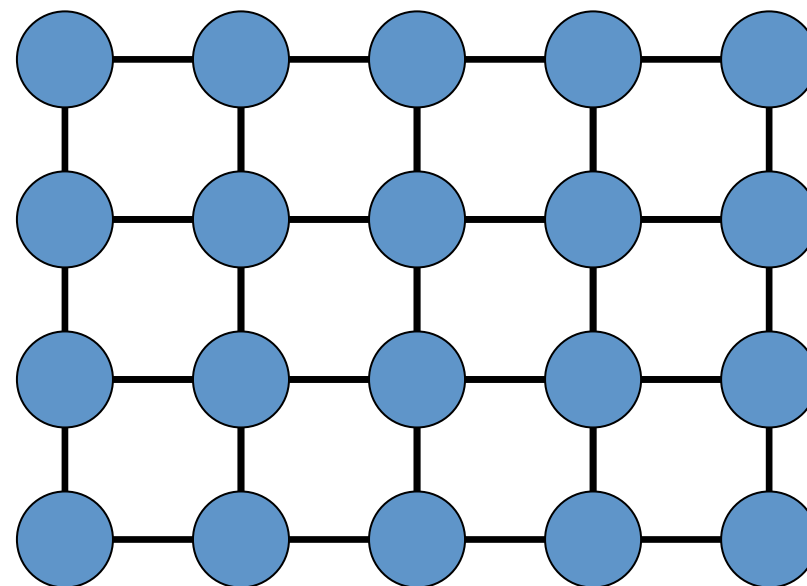
Object - white, Background - green/grey

Cost of a labelling  $f : V \rightarrow L$



Cost of same label high

Cost of different labels low



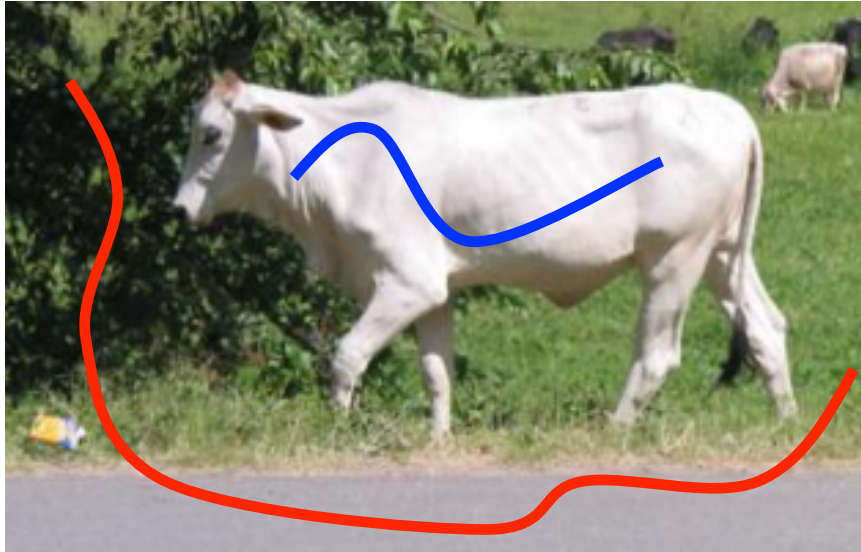
Graph  $G = (V, E)$

Per Edge Cost

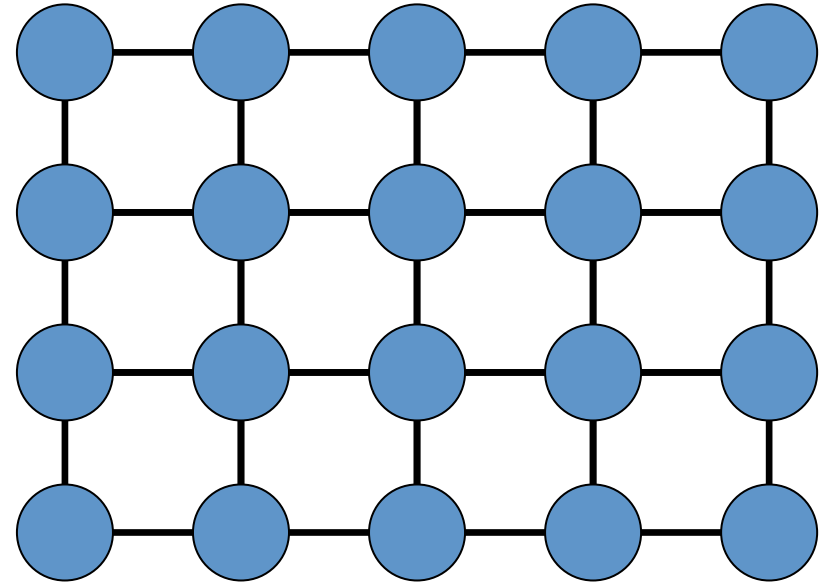
**PAIRWISE  
COST**

# A Computer Vision Application

## Binary Image Segmentation



Object - white, Background - green/grey



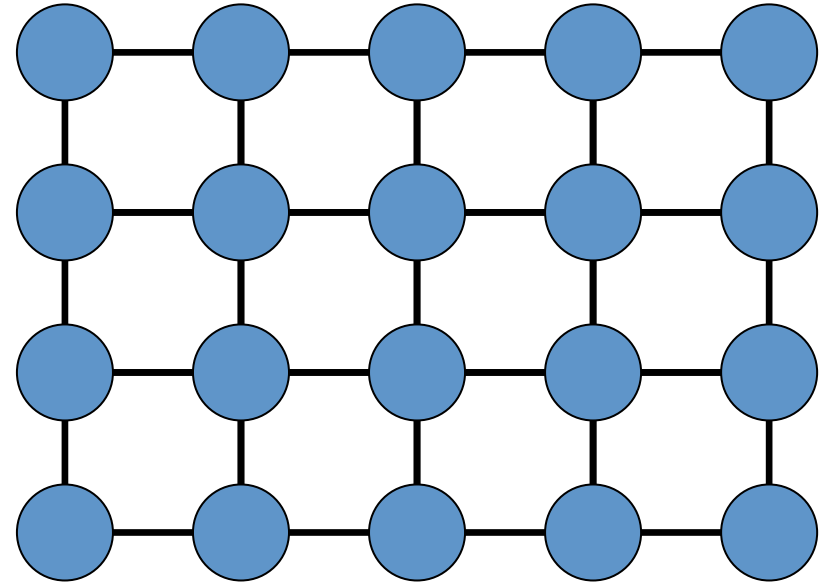
Graph  $G = (V, E)$

Problem: Find the labelling with minimum cost  $f^*$



# A Computer Vision Application

## Binary Image Segmentation

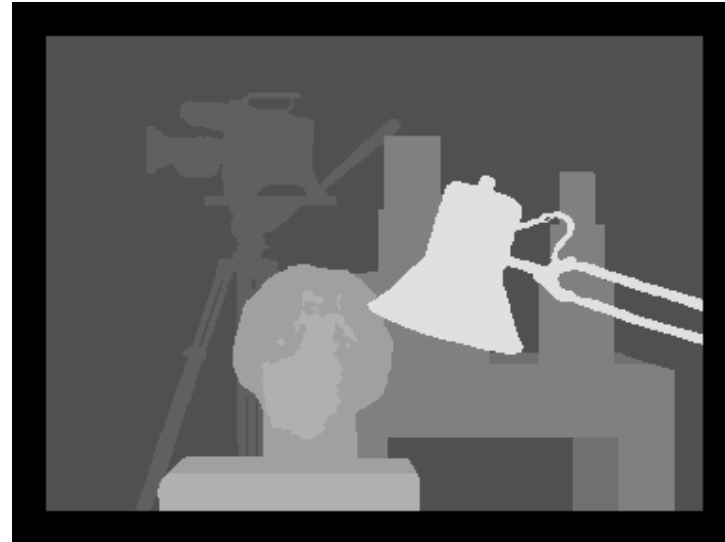
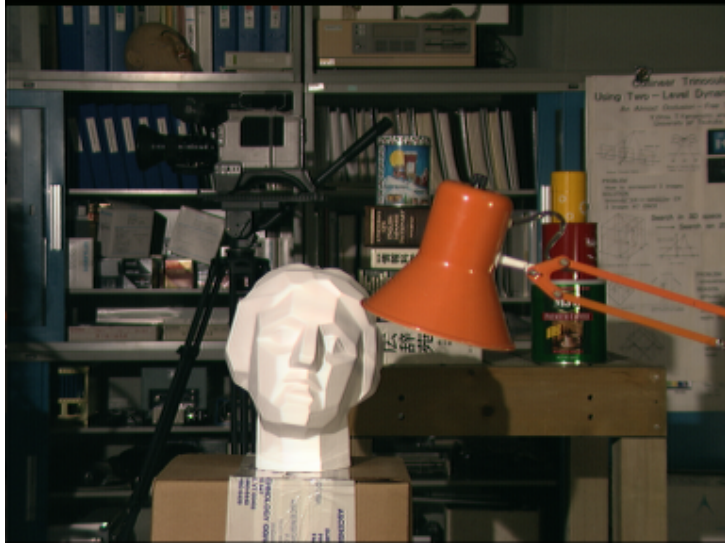


Graph  $G = (V, E)$

Problem: Find the labelling with minimum cost  $f^*$

# Another Computer Vision Application

## Stereo Correspondence



Disparity Map

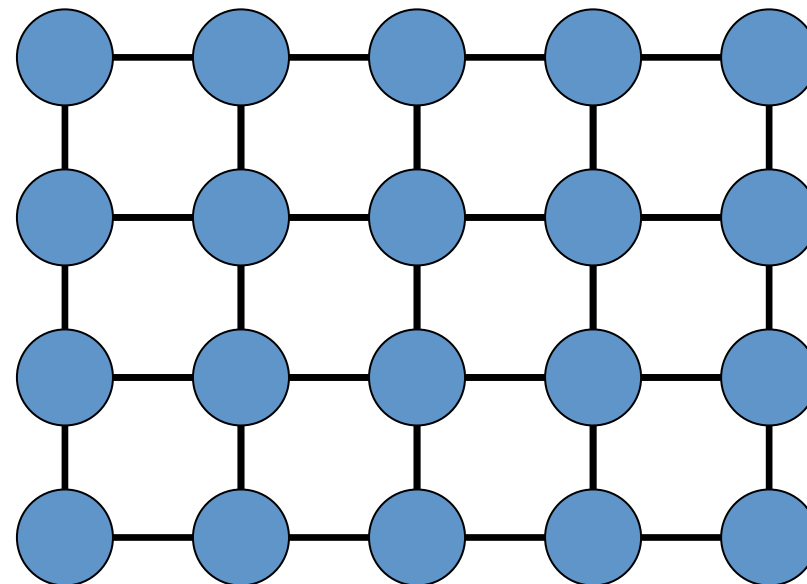
**How ?**

Minimizing a cost function



# Another Computer Vision Application

## Stereo Correspondence



Graph  $G = (V, E)$

Vertex corresponds to a pixel

Edges define grid graph

$L = \{\text{disparities}\}$

# Another Computer Vision Application

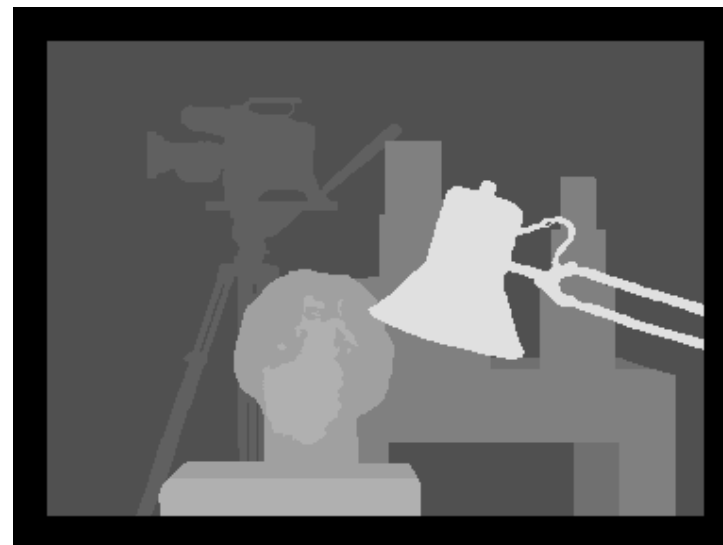
## Stereo Correspondence



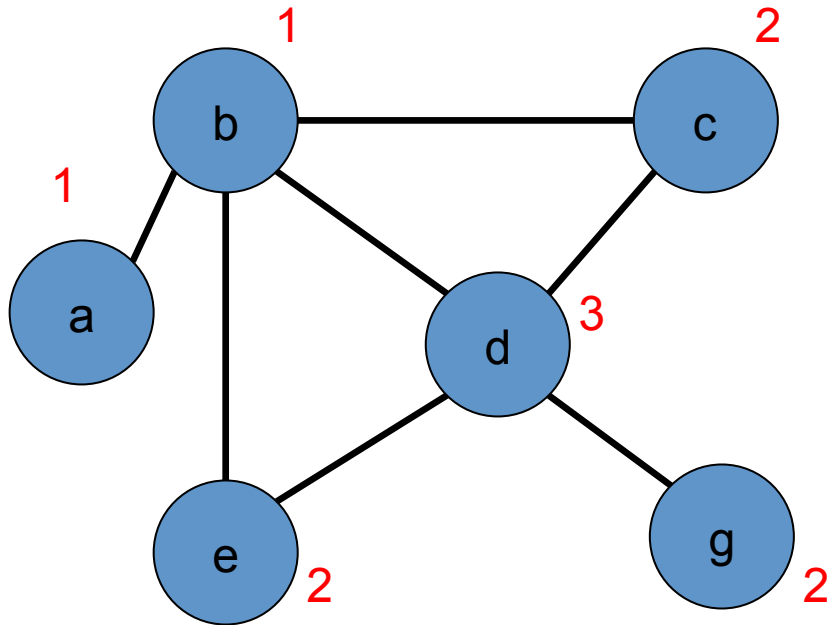
Cost of labelling  $f$  :

Unary cost + Pairwise Cost

Find minimum cost  $f^*$



# The General Problem



Graph  $G = (V, E)$

Discrete label set  $L = \{1, 2, \dots, h\}$

Assign a label to each vertex

$$f: V \rightarrow L$$

Cost of a labelling  $Q(f)$

Unary Cost

Pairwise Cost

Find  $f^* = \arg \min Q(f)$

# Overview

- Basics: problem formulation
  - Energy Function
  - MAP Estimation
  - Computing min-marginals
  - Reparameterization
- Solutions
  - Belief Propagation and related methods [Lecture 2]
  - Graph cuts [Lecture 3]