
Lecture 1

Introduction: Graphs and Probability

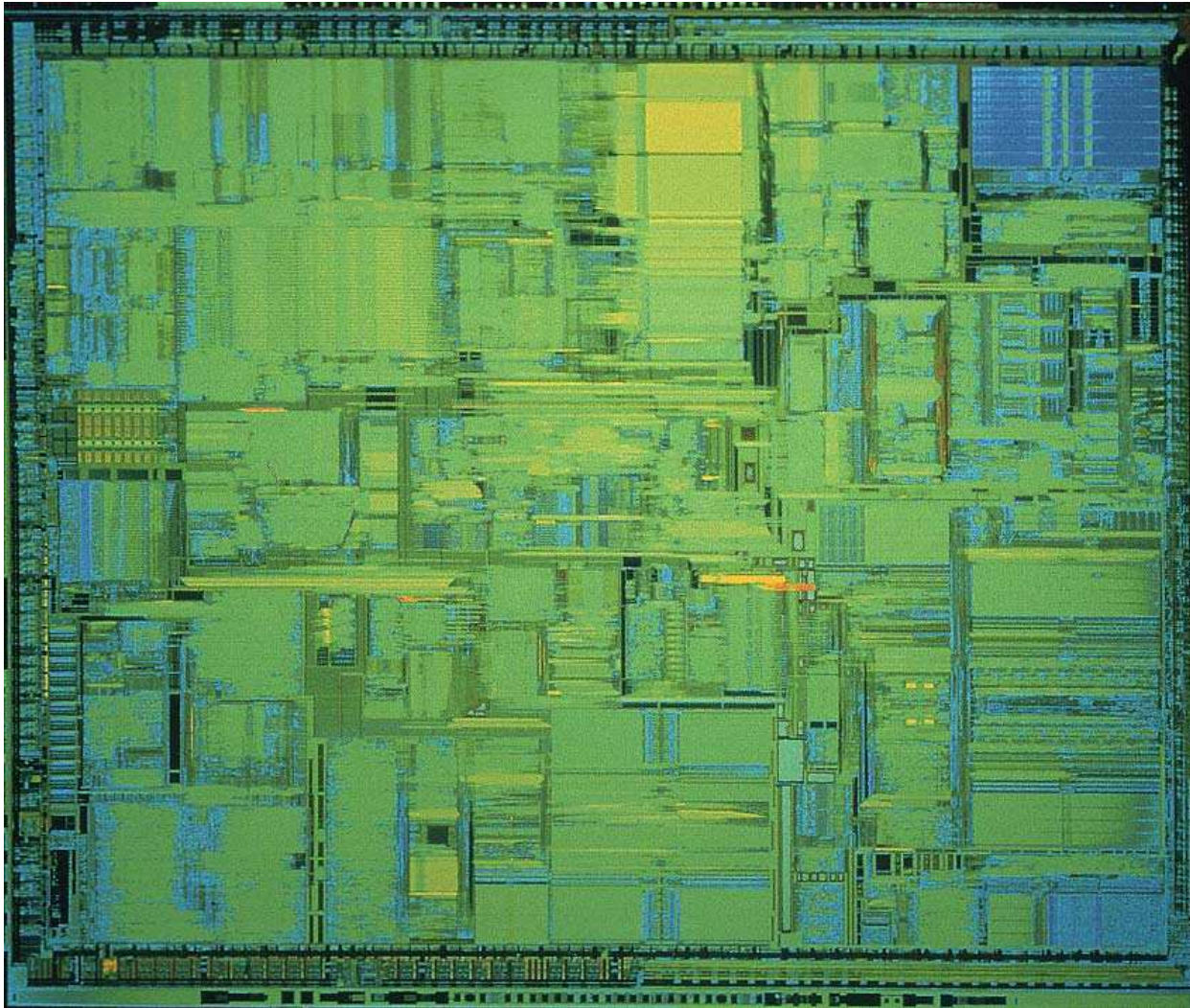


Overview

- **Graphical Models**
- Probabilities
- Modelling
- Modelling with Probabilistic Graphical Models



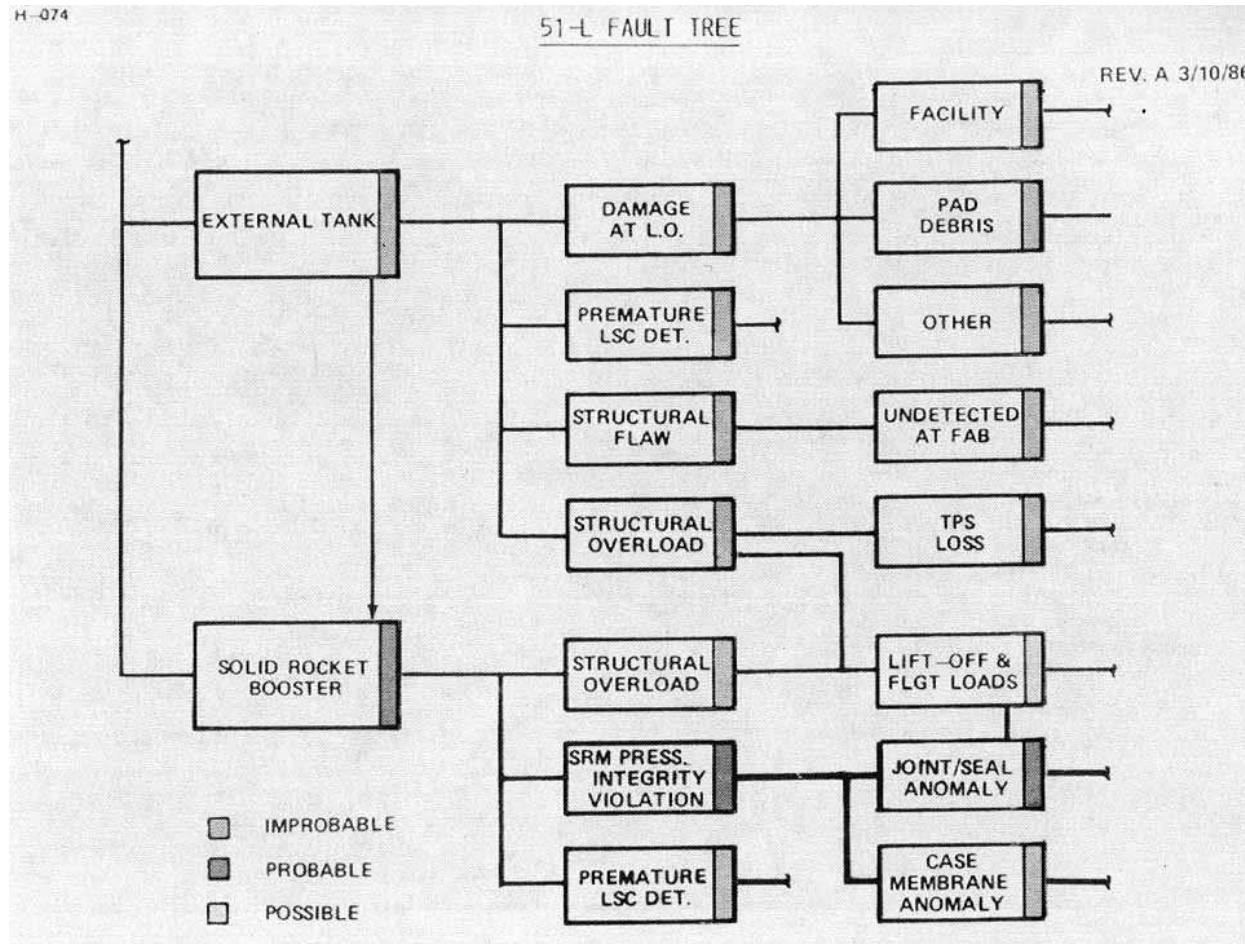
A Very Big Graphical Model



“Pentium 3” oil painting by Silicon Valley’s premier corporate artist Jens Gebhart.



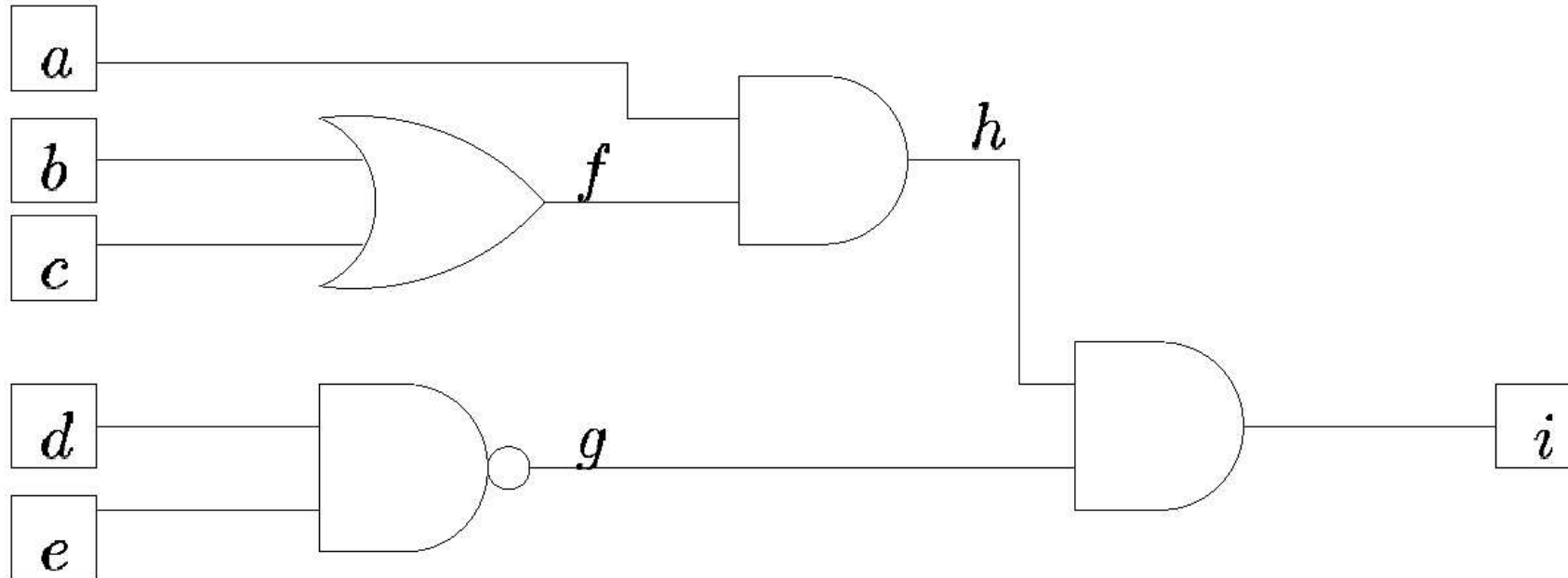
An Infamous Graphical Model



Part of a **fault tree** from Report of the PRESIDENTIAL COMMISSION on the Space Shuttle Challenger Accident, 1986.



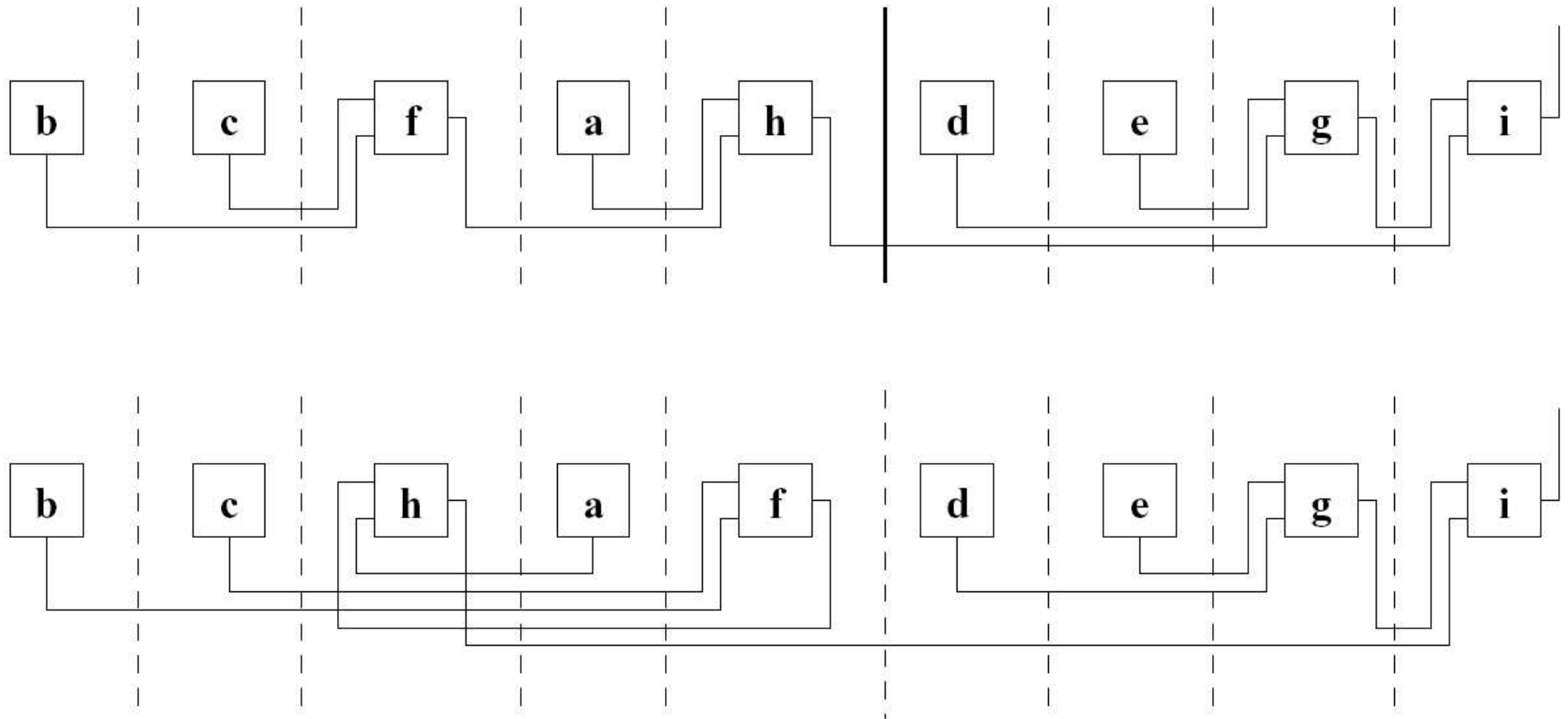
Constraint graphs from Circuits



$$(\bar{b} + f)(\bar{c} + f)(b + c + \bar{f})(d + g)(e + g)(\bar{d} + \bar{e} + \bar{g})(a + \bar{h})$$
$$(f + \bar{h})(\bar{a} + \bar{f} + h)(h + \bar{i})(g + \bar{i})(\bar{h} + \bar{g} + i)$$



Constraint Graphs from Circuits, linear layout



Two linear layouts.



A Microsoft Perspective



Gates responded ... Microsoft's competitive advantage ... was its expertise in 'Bayesian networks.' ... Is Gates onto something? Ask any other software executive about anything 'Bayesian' and you're liable to get a blank stare. Is this alien-sounding technology Microsoft's new secret weapon?

Los Angeles Times, October 28, 1996



Graphical Models

- ubiquitous in Computer Science
- directed, undirected, and hypergraphs graphs used;
- some applied examples:
 - computer networks (PCs, printers, hubs, ...)
 - telephone circuits
 - the Internet
 - circuit-level design for CPU chip
- some theoretical examples:
 - traveling salesman problem (TSP)
 - graph coloring
 - constraint graphs (e.g., for satisfiability)
 - hypergraph partitioning
- what is the *most studied* and *most experimented with*?



Graphical Models, cont.

- some model exact processes
- some model approximate processes:
 - expert systems' models
 - fault trees
 - models in image, speech, vision, etc.
- represent the structural aspects of a problem
- pictorial representations not necessary, just useful
- graphs focus on properties of binary relations; hypergraphs focus on n-ary relations
- we will cover some aspects of general theory useful for probabilistic graphs



Graphical Models: Basic Theory

The following definitions are given on MathWorld (<http://mathworld.wolfram.com>): graph, directed acyclic graph, forest, tree, degree, outdegree and indegree, clique, induced subgraph, and connected component.

Other definitions we will introduce:

- parents, ancestors, ancestral set
- children, descendants, descendant set
- cut-width (of elimination ordering)



Ancestors and Descendants

- Graph is a pair of sets (V, E) , for V the vertex set and E the edge set (a subset of all possible pairs from V).
- If $(x, y) \in E$, then x is a *parent* of y and y a *child* of x .
- The *transitive closure* of relation R is given by $R^* = I \cup R \cup R^2 \cup R^3 \cup \dots$

$$\text{ancestors}(S) \equiv \text{parents}^*(S) - S$$

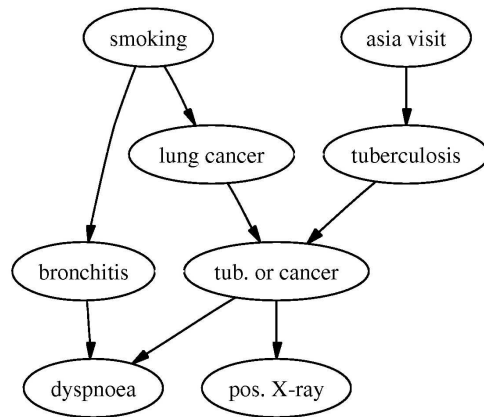
$$\text{descendants}(S) \equiv \text{children}^*(S) - S$$

$$S \text{ is an } \textit{ancestral set} \equiv \text{ancestors}(S) = \emptyset$$

$$S \text{ is a } \textit{descendant set} \equiv \text{descendants}(S) = \emptyset$$



Ancestors and Descendants, examples



ancestral sets $:: \{A\}, \{S\}, \{A, T\}, \{S, L, B\},$
 $\{A, S, T, L, B\}, \dots$

descendant sets $:: \{X\}, \{D\}, \{X, E, D\},$
 $\{T, L, X, E, D\}, \dots$

ancestors(X) $= \{A, T, E, L, S\}$

descendants(A) $= \{T, E, X, D\}$



Overview

- Graphical Models
- **Probabilities**
- Modelling
- Modelling with Probabilistic Graphical Models



Probability

Consider the probability of:

1. A strong virus alert will be announced for Windows in the next week.
 2. The Euro will go above US\$1.20 in 2004.
 3. The height of a Finnish male is 180–185cm.
 4. A Finnish male is tall.
- Probabilities must be for well-defined events: (4) is not, (1) possibly not.
 - Those for one-off unrepeatable events (1), (2) cannot be sampled.
 - Probabilities are always context dependent: (2) will vary during 2004, (3) changes when in a päiväkoti.
 - Continuous events need to be discretized, like (3).



Probability Operations

- Probability or probability density is a function over discrete or continuous variables in a domain.
- Operations and Concepts:

Conditioning	$p(X Y) = p(X,Y)/p(Y)$
Marginalization	$p(X) = \sum_Y p(X,Y)$
Factoring	$p(X,Y) = p(X)p(Y X)$
Independence	$X \perp\!\!\!\perp Y Z \equiv p(X Y,Z) = p(X Z)$ when $p(Y,Z) > 0$
Expectation	$\text{ExpectedValue}(U(X)) = \sum_X U(X)p(X)$
Maximization	$\text{Max}_d \text{ExpectedValue}(U(X,d))$

- Summation replaced by integration in continuous domains.
- Probability density is *always* a theoretical abstraction because the world and all measurements we take are discrete.



Constraint Satisfaction versus Probabilistic Reasoning

Correspondences roughly are as follows:

set X satisfies constraints	\leftrightarrow	$p(X) > 0$
set X fails constraints	\leftrightarrow	$p(X) = 0$
$\exists_X \wedge_i C_i(X)$	\leftrightarrow	$\sum_X \prod_i p_i(X)$
variable assignment	\leftrightarrow	conditioning
variable elimination	\leftrightarrow	marginalization
problem decomposition	\leftrightarrow	independence
local search	\leftrightarrow	Gibbs sampling

Similar mappings apply to optimization. Thus many relationships exist between graph algorithms for constraint satisfaction, optimization and probabilistic reasoning.



Three Kinds of Probabilities for the Bayesian

Proportions: e.g., probability the height of a Finnish male is 140–145cm

- true proportions about the world
- measurable but sometimes approximated
- Kolmogorov Axioms hold

Beliefs: e.g., probability Euro will go above US\$1.20 in 2004

- subjective beliefs, particular on one-off events
- not practically measurable, realm of Decision Theory
- Cox's Axioms, Dutch books or other schemes justify

Beliefs about Proportions: e.g., probability density function on proportions for rolls of die for biased dice on Teemu's desk

- subjective beliefs about true proportions
- not practically measurable
- realm of Bayesian Statistics



Assumptions About Probability as Belief

Standard frameworks for justifying the use of probabilities to model beliefs (i.e., Bayesian reasoning) make the following (often hidden) **assumptions**:

- you have unlimited computational power
- you are a single agent seeking to optimize performance in directly measurable sense
- you are using a model family that includes a sufficiently close approximation to the “truth”

These become **dangerous** when:

- seeking to model “objective” reasoning
- using simple linear models for everything
- using an old CPU with low power and memory



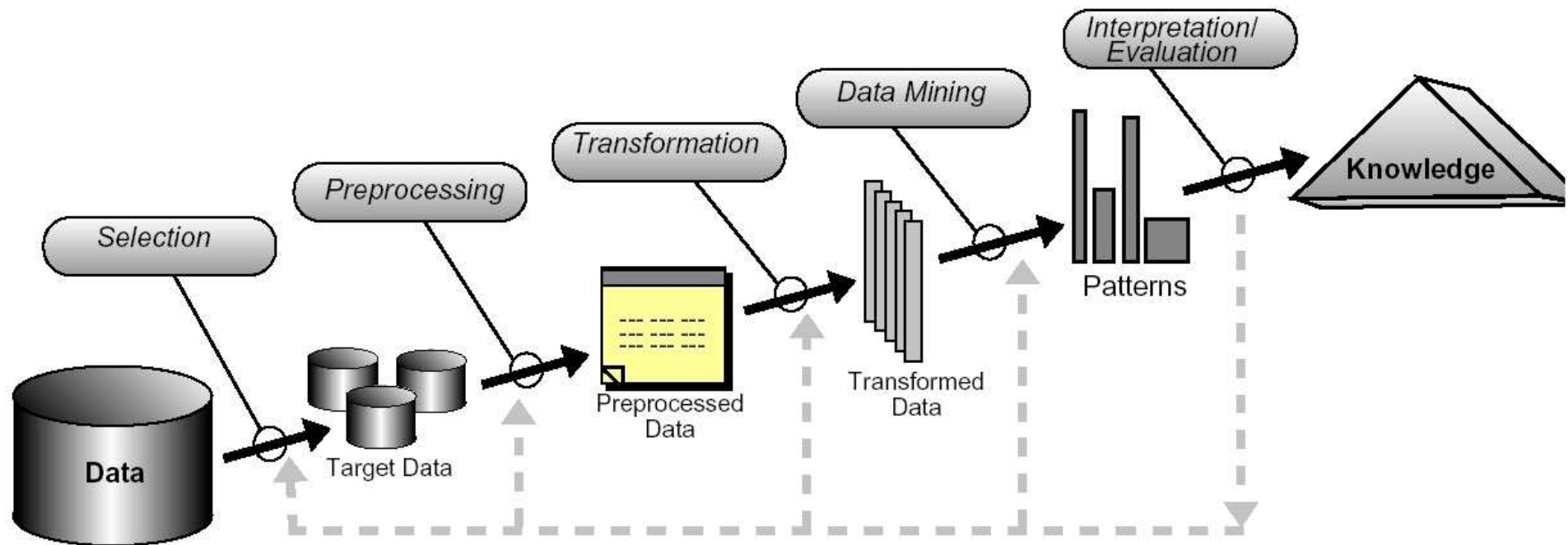
Overview

- Graphical Models
- Probabilities
- **Modelling**
- Modelling with Probabilistic Graphical Models



The Knowledge Discovery Process

Is a process of iterative refinement. But represented here without a “model”.

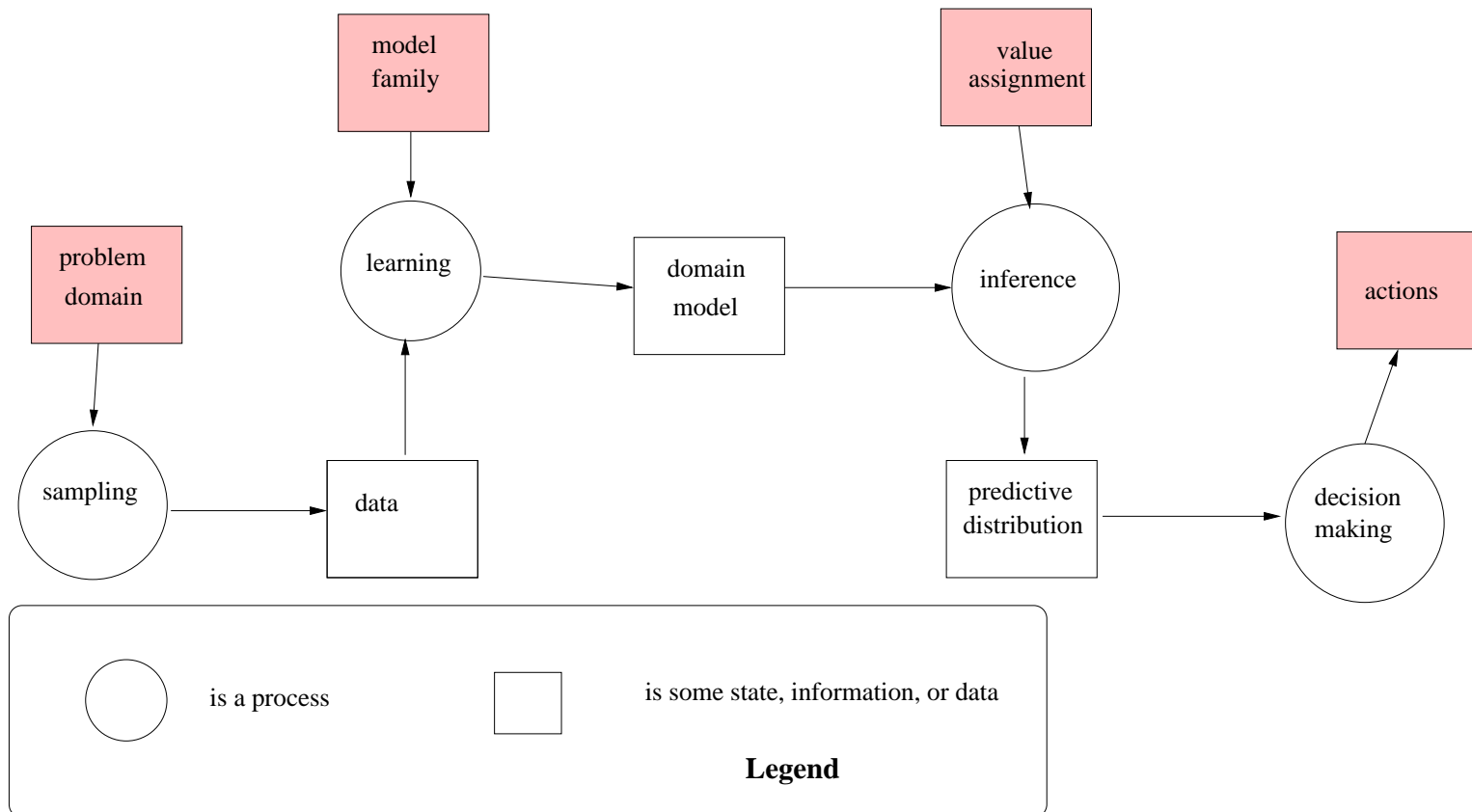


From Fayyad, Piatetsky-Shapiro & Smyth (1996)



The Modelling Process

Is a process of iterative refinement (though here static). A model represents a *generative* or *functional* model for the problem elements.



From Tirri's class notes for Three Concepts: Probability



Modelling Objectivity

Question: *The Royal Society wants to make a report about whether the new AIDS drug KylieKleptron cause dangerous side effects?*

- Their statement should not be subjective.
- Their statement should not have to be retracted in 12 months due to additional evidence.
- Their statement should reflect the considered opinion of a variety of different experts.

Objective reasoning using probabilities requires different arguments and methods beyond simple Bayesian probabilistic reasoning and is not covered here.



Modelling Dimensions: Tradeoffs / Art

- More data allows for better calibration of your model, but may be expensive or costly.
- Using more complex models might be closer to “truth”, but might cause more computation.
- Human experts can often provide advise, but are fallible.
- Sometimes, just the choice of variables to use is the critical decision.
- Don't mess with the data! Model it.



Overview

- Graphical Models
- Probabilities
- Modelling
- **Modelling with Probabilistic Graphical Models**

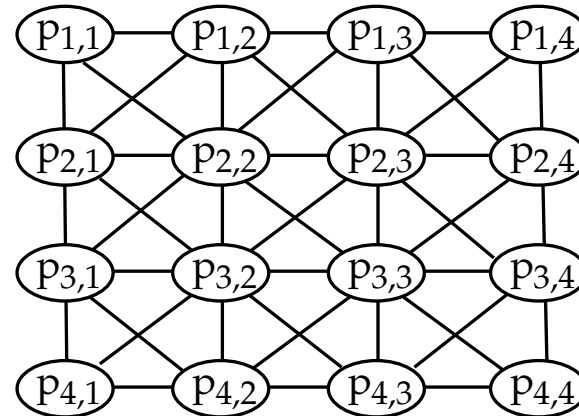
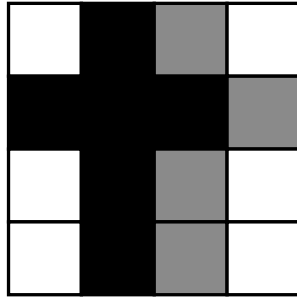


Probabilistic Graphical Models

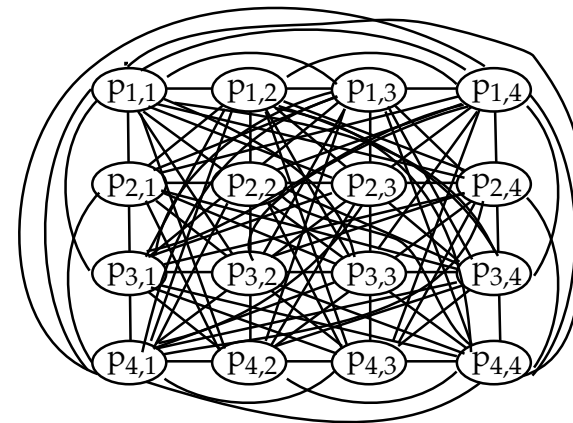
- different kinds of graphs for different kinds of information:
 - causes
 - dependencies, direct influences, . . .
 - actions, states, observables, . . .
- probabilistic relations are associated with graphs i.e., represents a set of probability equations
 - Gaussians, linear models
 - discrete distributions
 - neural nets
- standard graph operations (variable elimination, decomposition, etc.) apply



Simple Image Models



Simple 4 × 4 image. Top graph says all pixels influenced only by their neighbour's values. Bottom graph says everyone influences everyone else.



Next Week

- Review Shachter's notes on "Examples without Numbers"
- Review Scheines' online tutorial on "d-separation" (but don't spend too much time, we won't use this).
- Check out suggested project topics (should be online by Tuesday next) and consider a topic of your own choosing.

