电子科技大学
University of Electronic Science and Technology of China

Probabilistic Graphical Models-chapter1&2

Xinwu Chen

Data Mining Lab, Big Data Research Center, UESTC
Email：junmshao@uestc.edu.cn
http://staff.uestc.edu.cn/shaojunming

Warning !

- The stories, formulas definitions in my speech you may have heard thousand of times

- The 1st and 2nd chapter is just an preface of probabilistic graphical model

Contents

数据挖掘实验室
Data Mining Lab

➢1 Introduction

- 1.1 Motivation
- 1.2 Structured Probabilistic Models
- 1.3 Overview and Roadmap(Ø)
- 1.4 Historical Notes(Ø)

➢2 Foundations

- 2.1 Probability Theory
- 2.2 Graphs

1.1 Motivation

数据挖掘实验室
Data Mining Lab

- The true state of world is fundamental uncertainty
- We can’t figure out what makes the stock market fall

- To obtain meaningful conclusions, we need to reason not just about what is possible, but also about what is probable

1.1 Motivation

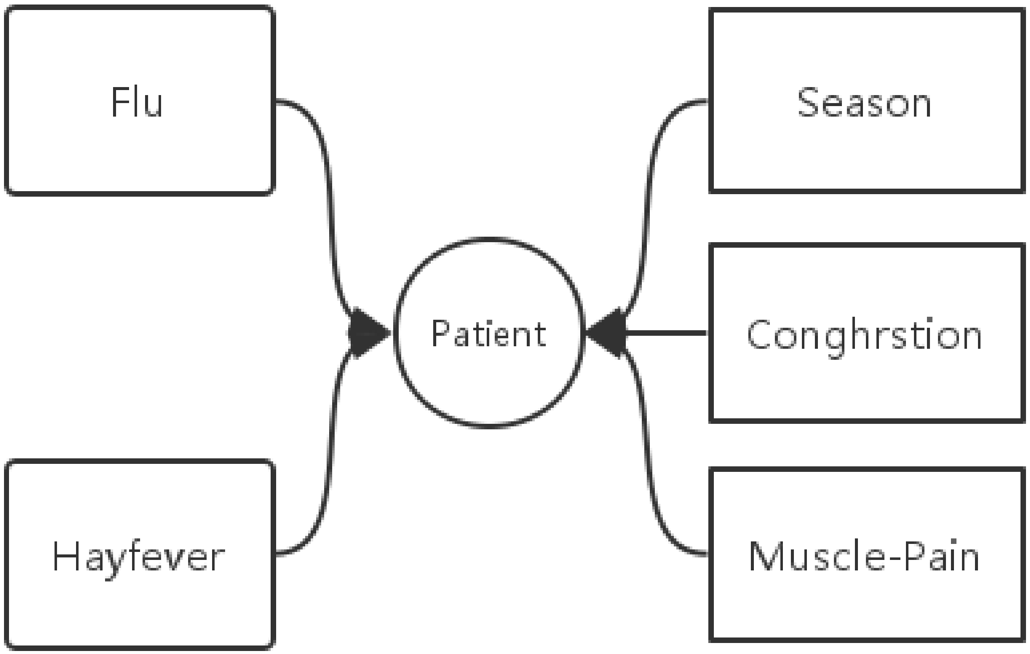
➢We want

- Take the available information and reach conclusions
- Develop a general suite of algorithms that apply any model within a broad class
- Solve the problem of the noisy etc
- Good for your study

1.2 Structured Probabilistic Models

数据挖掘实验室
Data Mining Lab

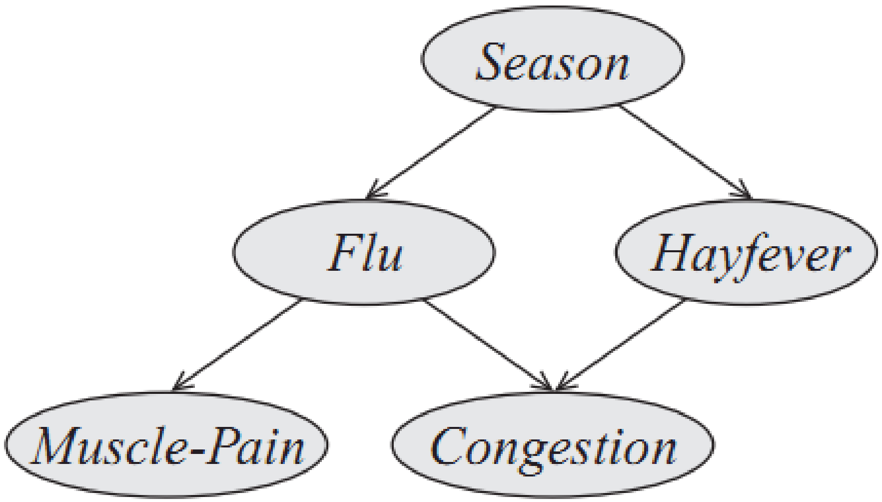
➢Example



2*2*2*2*4=64

1.2 Structured Probabilistic Models

➢Example



Independencies

$$(F \perp H \mid S)$$
$$(C \perp S \mid F, H)$$
$$(M \perp H, C \mid F)$$
$$(M \perp C \mid F)$$

Factorization

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

➢Example

“spring, no flu, hayfever, inus congestion, muscle pain”

- P(Season = spring)*
- P(Flu = false | Season = spring)*
- P(Hayfever = true | Season = spring)*
- P(Congestion = true | Hayfever = true; Flu = false)*
- P(Muscle Pain = true | Flu = false)

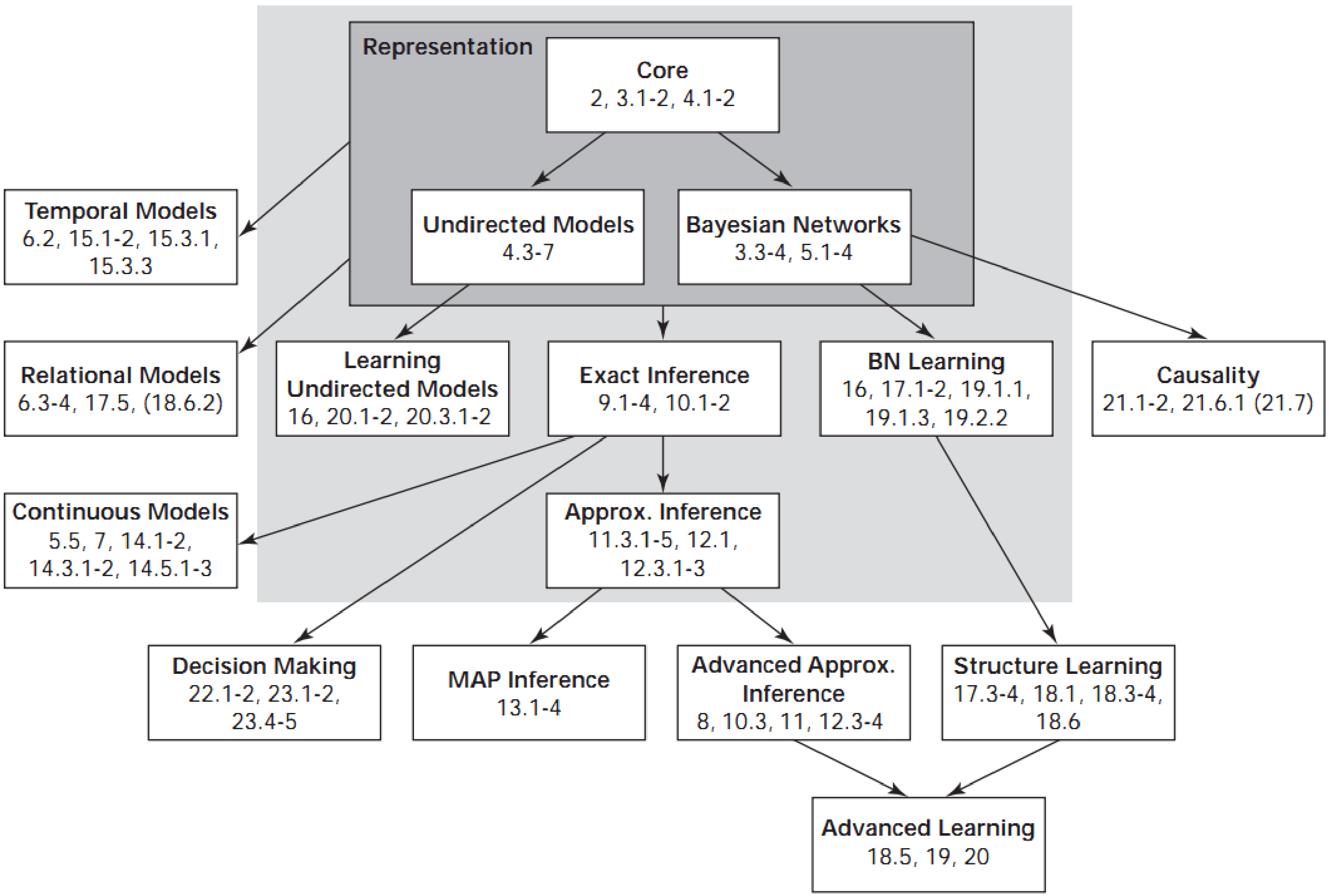
✓ 3+4+4+4+2 = 17

➢ **The property that variables tend to interact directly only with very few others**

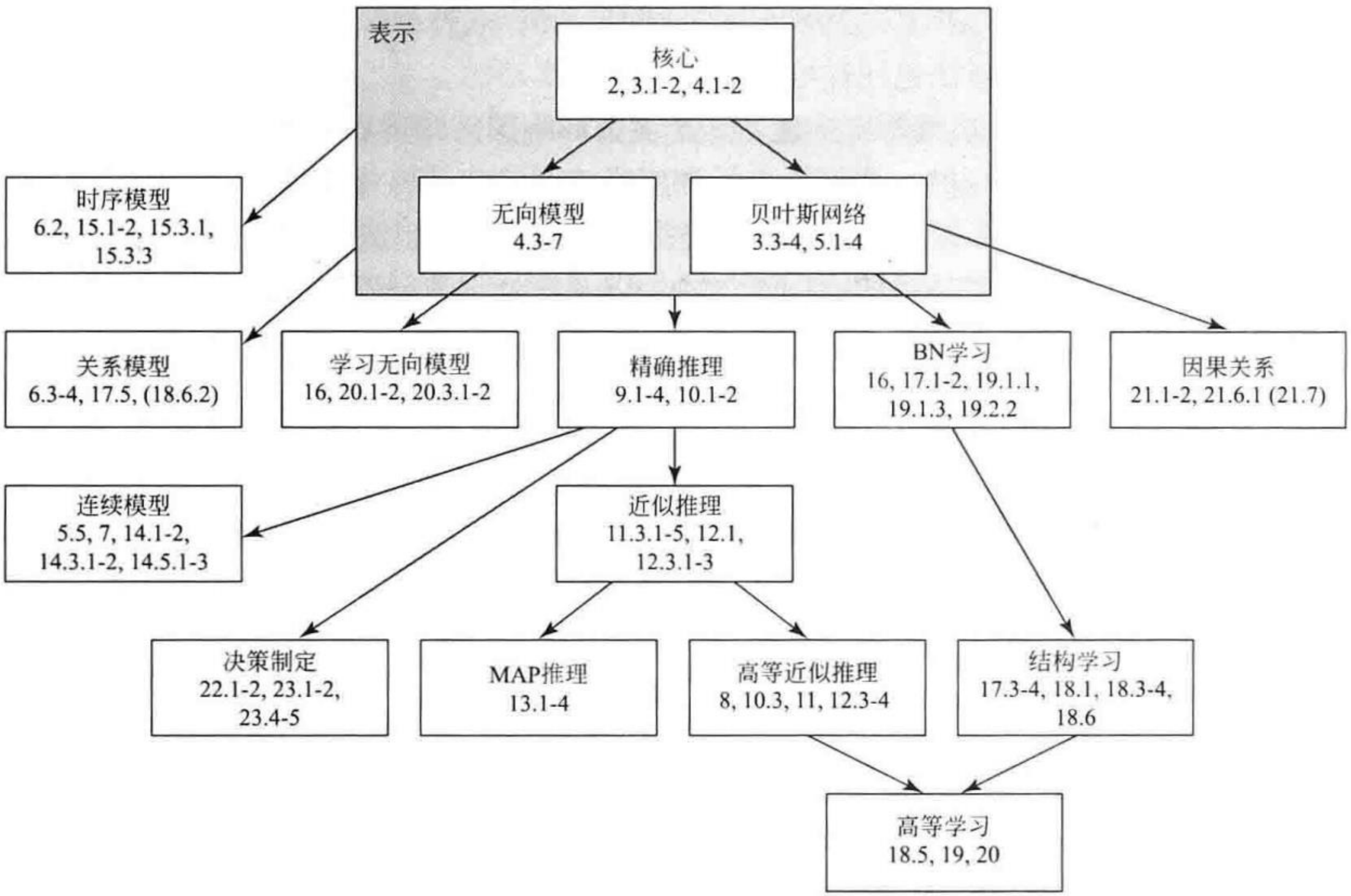
- **1.It often allows the distribution to be written down tractably**
- **2. The same structure often also allows the distribution to be used effectively for inference — answering queries using the distribution as our model of the world**
- **3. This framework facilitates the effective construction of these models**

1.2 Structured Probabilistic Models

数据挖掘实验室
Data Mining Lab



1.2 Structured Probabilistic Models



➢ Probability Distribution

- **Event Spaces** Ω $\qquad \Omega = \{1,2,3,4,5,6\}$

1	2	3	4	5	6

- **Measurable events** S
  - $\phi \in S, \Omega \in S$
  - $\forall \alpha, \beta \in S, \alpha \cup \beta \in S$
  - $\alpha \in S, \Omega - \alpha \in S$

$$S = \{\phi, \{1,3,5\}, \{2,4,6\}, \Omega\}$$

$$S = \{\phi, \Omega\}$$

1	2	3	4	5	6

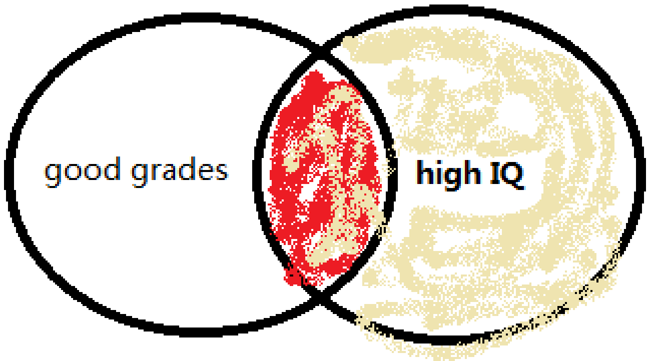
- **Probability distribution** (Ω, S) $S \to R$
  - $\forall \alpha \in S, P(\alpha) \geq 0$
  - $P(\Omega) = 1$
  - $\forall \alpha, \beta \in S, \alpha \cap \beta = \phi \Rightarrow P(\alpha \cup \beta) = P(\alpha) + P(\beta)$

2.1 Probability Theory

- **Conditional Probability** $P(\alpha \mid \beta) = \dfrac{P(\alpha \cap \beta)}{P(\beta)}$

  α: students with good grades

  β: students with high IQ



- **Chain rule**

  $P(\alpha_1 \cap ... \cap \alpha_k) = P(\alpha_1)P(\alpha_2 \mid \alpha_1)...P(\alpha_k \mid \alpha_1 \cap ... \alpha_{k-1})$

- **Random Variable** $\Omega \rightarrow \text{value}$

  $A = GradeA, B = high$

- **Joint Distribution** $\chi = \{X_1, ... X_k\} \rightarrow P(\chi)$
- **Marginal Distribution** $P(X) = \sum_y P(X, y)$
- **Bayes' rule**

$$P(\alpha \mid \beta) = \frac{P(\beta \mid \alpha)P(\alpha)}{P(\beta)}.$$

		Intelligence low	*Intelligence* high	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

➢Independence and Conditional Independence

- **Independence**

$$P(\alpha \cap \beta) = P(\alpha)P(\beta), P(\alpha | \beta) = \frac{P(\alpha \cap \beta)}{P(\beta)} = P(\alpha)$$

$$P(\alpha | \beta) \neq P(\alpha)$$

$$P(\alpha | \beta) = P(\alpha) \| P(\beta) = 0, P \models (\alpha \perp \beta)$$

$$(\alpha \perp \beta) \Leftrightarrow (\beta \perp \alpha)$$

- **Conditional Independence**
- **P(MIT|Stanford,GradeA) = P(MIT|GradeA)**

$$(\alpha \perp \beta | \gamma) \Leftrightarrow P(\alpha \cap \beta | \gamma) = P(\alpha | \gamma)P(\beta | \gamma)$$

2.1 Probability Theory

➢Independence properties

- **Symmetry** $(X \perp Y \mid Z) \Leftrightarrow (Y \perp X \mid Z)$
- **Decomposition** $(X \perp Y, W \mid Z) \Rightarrow (X \perp Y \mid Z)$
- **Weak union** $(X \perp Y, W \mid Z) \Rightarrow (X \perp Y \mid W, Z)$
- **Contraction** $(X \perp W \mid Z, Y) \,\&\, (X \perp Y \mid Z) \Rightarrow (X \perp Y, W \mid Z)$
- **Intersection** $(X \perp Y \mid Z, W) \,\&\, (X \perp W \mid Z, Y) \Rightarrow (X \perp Y, W \mid Z)$

2.1 Probability Theory

➢Querying a Distribution

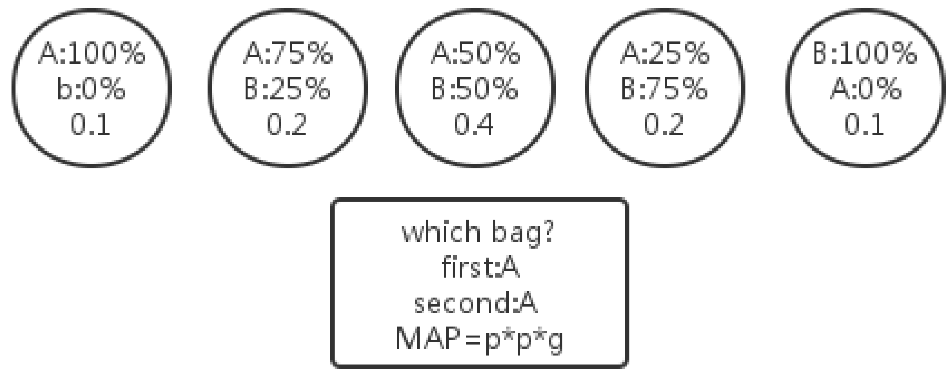
- **Probability Queries**

  $P(Y|E = e)$

- **MAP Queries**

$$\mathrm{MAP}(\boldsymbol{W} \mid \boldsymbol{e}) = \arg\max_{\boldsymbol{w}} P(\boldsymbol{w}, \boldsymbol{e})$$

- **Marginal MAP Queries**



2.1 Probability Theory

➢probability density function(PDF)

$$P(a \le X \le b) = \int_a^b p(x)dx.$$

➢Joint Density Functions

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} p(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

➢Conditional Density Functions

$$P(Y \mid x) = \lim_{\epsilon \longrightarrow 0} P(Y \mid x - \epsilon \le X \le x + \epsilon).$$

➢Expectation

$$\mathbb{E}_P[X] = \sum_x x \cdot P(x).$$

Variance

$$\mathbb{V}ar_P[X] = \mathbb{E}_P\left[(X - \mathbb{E}_P[X])^2\right].$$

➢ Chebyshev inequality

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{D[X]}{\varepsilon^2}$$

$$\varepsilon = k\sigma, P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

Indicator function $I_{\{A\}} = 1, A = true$

$$E[I_{\{A\}}] = P(A)$$

$$I_{\{|x-a| \geq b\}} \leq \frac{(x-a)^2}{b^2}$$

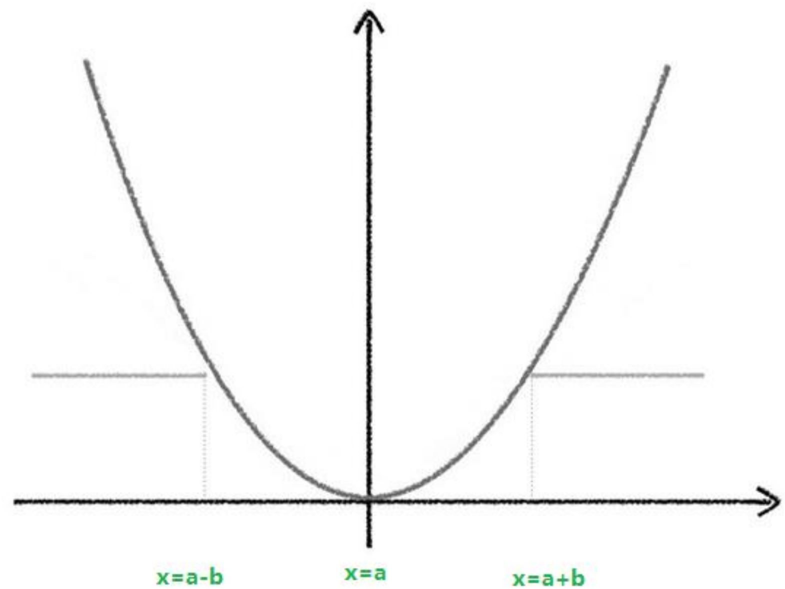
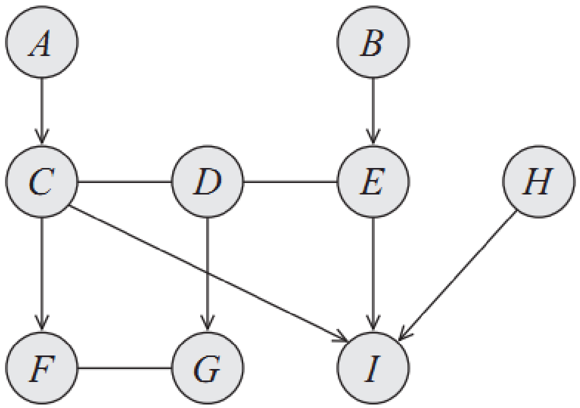


Image courtesy of Jim Pitman, Omid Solari and Xin Wang.

2.2 Graphs

数据挖掘实验室
Data Mining Lab

- A graph is a data structure K consisting of a set of nodes and a set of edges
- Node: X = {X1; ...;Xn}
- Edge:



- undirected edge Xi—Xj

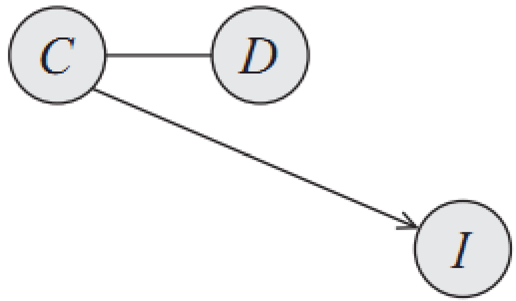
  Xi is a neighbor of Xj

- directed edge Xi -> Xj

  Xj is the child of Xi Xi is the parent of Xj

Given a graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$, *its* undirected version *is a graph* $\mathcal{H} = (\mathcal{X}, \mathcal{E}')$ *where* $\mathcal{E}' = \{X—Y : X \rightleftharpoons Y \in \mathcal{E}\}$. ■

➢2.2.2 Subgraphs



Let $\mathcal{K} = (\mathcal{X}, \mathcal{E})$, *and let* $\boldsymbol{X} \subset \mathcal{X}$. *We define the* induced subgraph $\mathcal{K}[\boldsymbol{X}]$ *to be the graph* $(\boldsymbol{X}, \mathcal{E}')$ *where* $\mathcal{E}'$ *are all the edges* $X \rightleftharpoons Y \in \mathcal{E}$ *such that* $X, Y \in \boldsymbol{X}$. ■

A subgraph over $\boldsymbol{X}$ *is* complete *if every two nodes in* $\boldsymbol{X}$ *are connected by some edge. The set* $\boldsymbol{X}$ *is often called a* clique; *we say that a clique* $\boldsymbol{X}$ *is* maximal *if for any superset of nodes* $\boldsymbol{Y} \supset \boldsymbol{X}$, $\boldsymbol{Y}$ *is not a clique.* ■

2.2 Graphs

➢2.2.2 Subgraphs

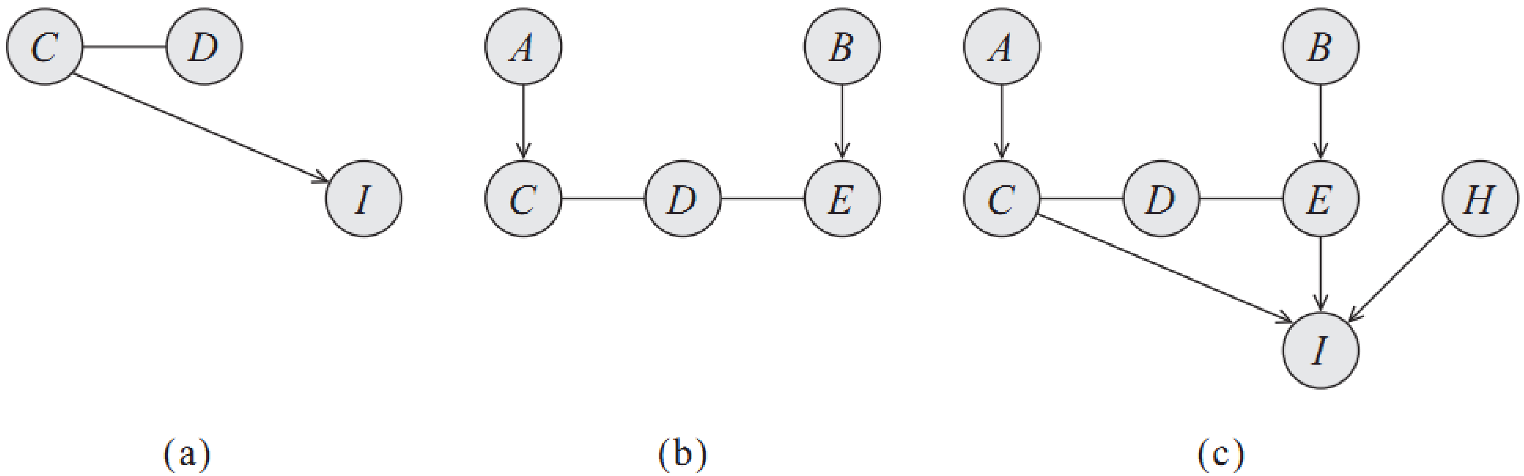
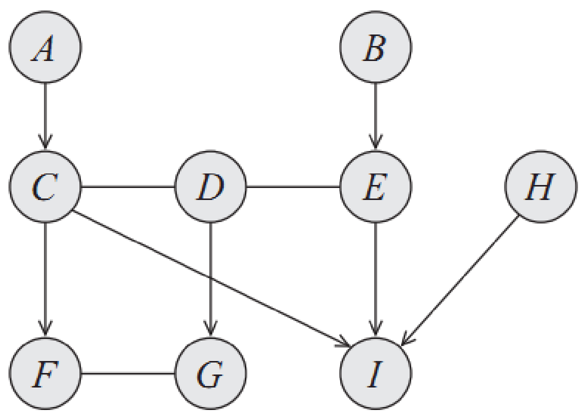


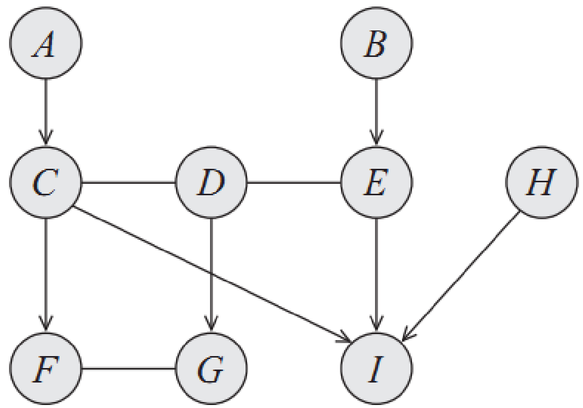
Figure 2.4 Induced graphs and their upward closure: (a) The induced subgraph $\mathcal{K}[C, D, I]$. (b) The upwardly closed subgraph $\mathcal{K}^{+}[C]$. (c) The upwardly closed subgraph $\mathcal{K}^{+}[C, D, I]$.



数据挖掘实验室
Data Mining Lab

➢2.2.3 Paths and Trails

- **Path**
- **Trail($X_i \rightleftharpoons X_{i+1}$)**



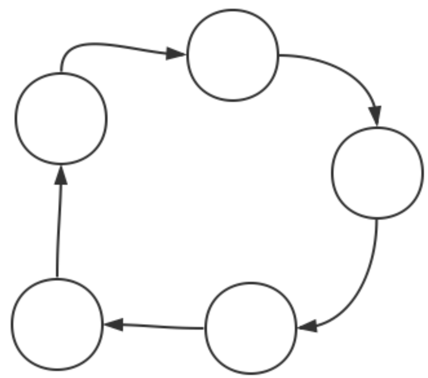
- **A graph is connected if for every Xi, Xj there is a trail between Xi and Xj**
- **Ancestors X**
- **Topological ordering**

Let $\mathcal{G} = (\mathcal{X}, \mathcal{E})$ be a graph. An ordering of the nodes $X_1, \ldots, X_n$ is a topological ordering *relative to $\mathcal{K}$ if, whenever we have $X_i \rightarrow X_j \in \mathcal{E}$, then $i < j$.* ■

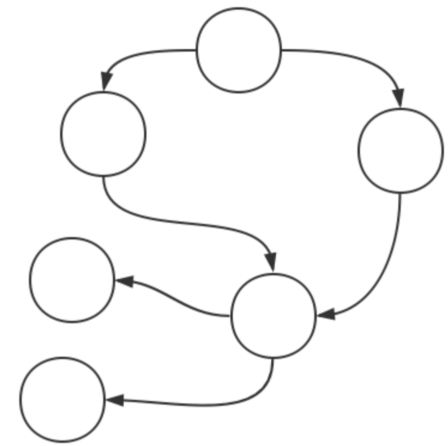
2.2 Graphs

➢2.2.4 Cycles and Loops

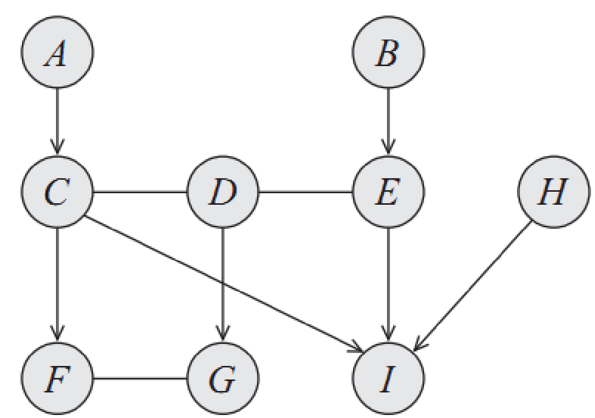
- Cycle



- Directed acyclic graph (DAG)
- [Bayes]

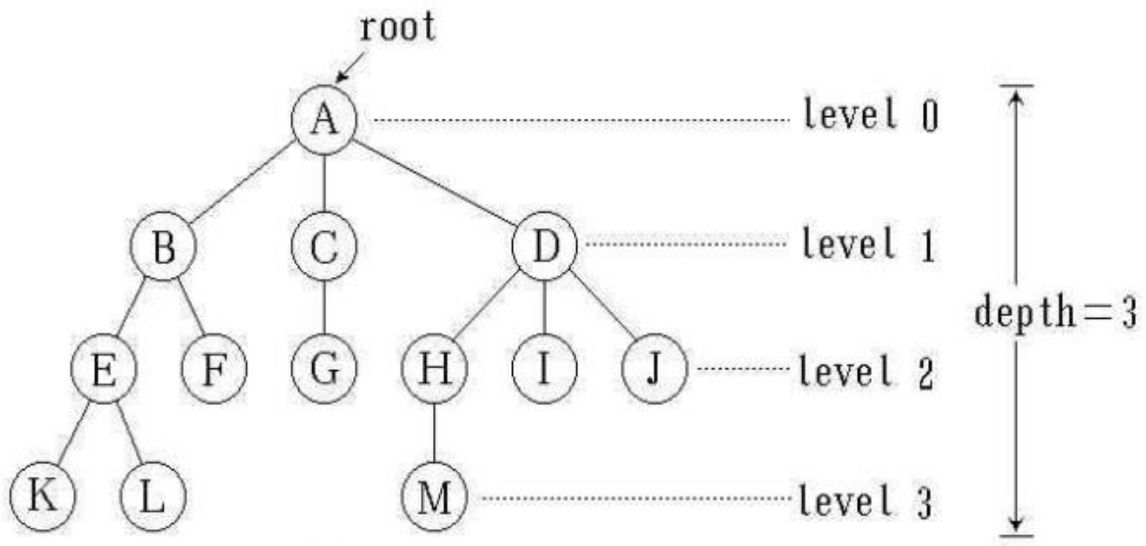


- Partially directed acyclic graph (PDAG)
- (Chain graph)
- {A}, {B}, {C;D;E}, {F;G}, {H}, {I}

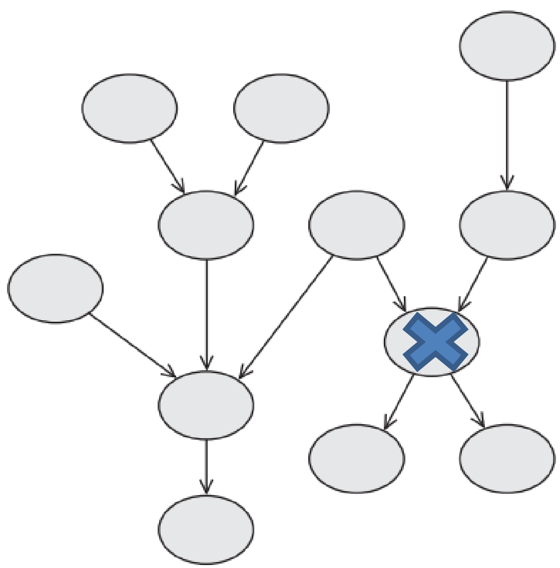


➢2.2.4 Cycles and Loops

- Tree
- Forest



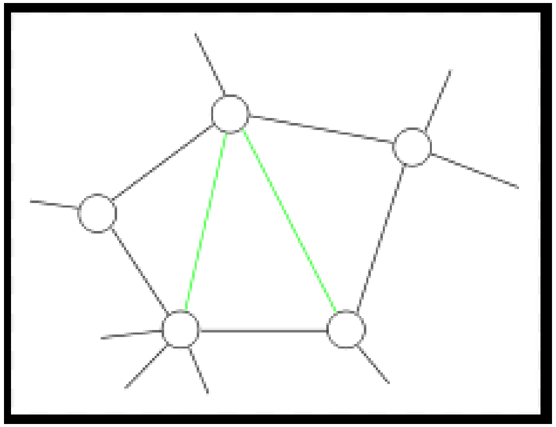
- polytree



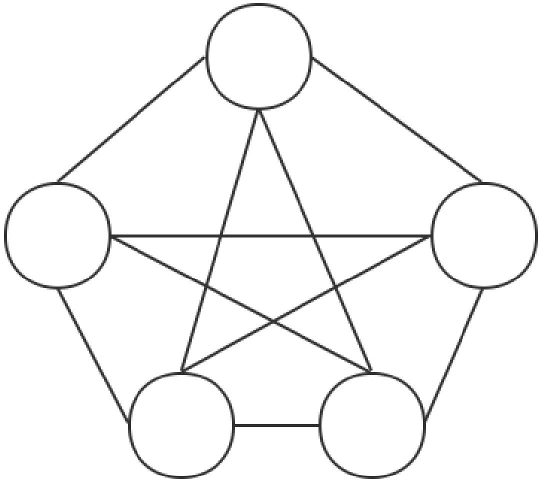
2.2 Graphs

➢2.2.4 Cycles and Loops

- Chord



- Triangulated
- chordal graph



Thanks

Xinwu Chen
University of Electronic Science and Technology of China
chen_xin_wu@qq.com