Intelligent Robotics

Graphs and Bayes Nets[1]

Mohan Sridharan
School of Computer Science
University of Birmingham, UK

October 8, 2019

[1] Based on material from "Probabilistic Graphical Models: Principles and Techniques" by Koller and Friedman

Lecture Outline

- **Graphs**:
  - Nodes and edges.
  - Paths and trails.
  - Cycles and loops.
- **Bayes nets**:
  - Basic concepts.
  - Reasoning patterns.
  - Markov networks.

Nodes and Edges 1

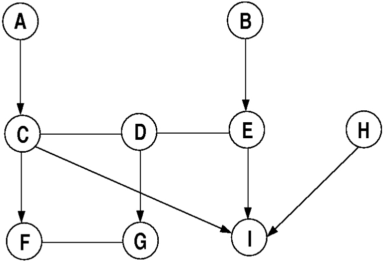
- A *Graph* is a data structure consisting of a set of nodes and edges:

$$\mathcal{K} = (\mathcal{X}, \mathcal{E}); \quad \mathcal{X} = \{X_1, \ldots, X_n\}$$

- Edges can be *directed*: $X_i \rightarrow X_j$ or *undirected*: X_i–X_j.
- Directed graph $\mathcal{G}$: all edges are directed.
- Undirected graph $\mathcal{H}$: all edges are undirected.

Nodes and Edges 2

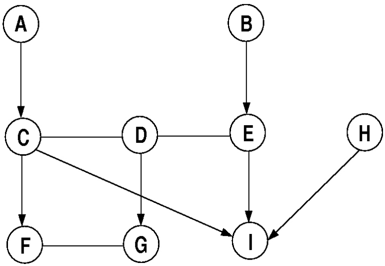
- If $X_i \rightarrow X_j \in \mathcal{E}$, X_i is the parent and X_j is the child.
- If $X_i – X_j \in \mathcal{E}$, X_i and X_j are neighbors.
- $X \rightleftharpoons Y$ are adjacent.
- $Boundary_X = Pa_X \cup Nb_X$.



- Subgraph over **X** is complete if every two nodes in **X** are connected by an edge. Also called a *clique*.
- Clique **X** is maximal is any **Y** $\supset$ **X** is not a clique.

Paths and Trails 1

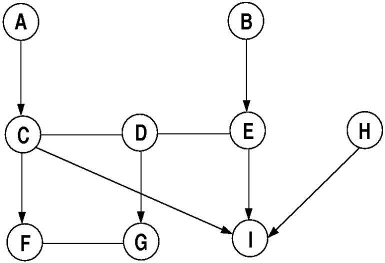
- $X_1, \ldots, X_k$ form a path in graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$ if for every $i = 1, \ldots, k-1$, either $X_i \rightarrow X_{i+1}$ or $X_i – X_{i+1}$.
- Directed path: for at least one i, $X_i \rightarrow X_{i+1}$.



- $X_1, \ldots, X_k$ form a trail in graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$ if for every $i = 1, \ldots, k-1$, $X_i \rightleftharpoons X_{i+1}$.

Paths and Trails 2

- A graph is connected if there is a trail between every X_i and X_j.
- If there exists a directed path $X_1, \ldots, X_k$ in $\mathcal{K} = (\mathcal{X}, \mathcal{E})$ with $X_1 = X, X_k = Y$, X is an ancestor of Y and Y is a descendant of X.



- $Descendants_A = ?$, $NonDescendants_B = ?$
- $X_1, \ldots, X_k$ form a topological ordering in graph $\mathcal{K} = (\mathcal{X}, \mathcal{E})$ if for every $X_i \rightarrow X_j \in \mathcal{E}$, we have $i < j$.

Cycles and Loops 1

- A cycle in $\mathcal{K}$ is a directed path $X_1, \ldots, X_k$ where $X_1 = X_k$.
- A graph is acyclic if it contains no cycles. *Bayesian Network* is a directed acyclic graph (DAG).
- A loop in $\mathcal{K}$ is a trail $X_1, \ldots, X_k$ where $X_1 = X_k$.
- A graph is singly connected if it has no loops.

Cycles and Loops 2 (Extra)

- Polytree: singly-connected directed graph.
- Leaf: node in singly-connected graph with exactly one adjacent node.
- Forest: singly-connected undirected graph. Also, directed graph where each node has at most one parent.
- Tree: connected forest.

Lecture Outline

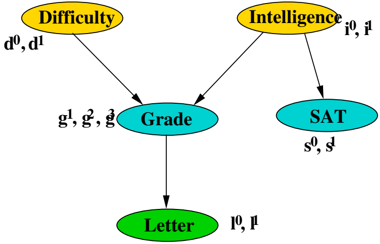
- **Graphs**:
  - Nodes and edges.
  - Paths and trails.
  - Cycles and loops.
- **Bayes nets**:
  - Basic concepts.
  - Reasoning patterns.
  - Markov networks.

Bayesian Network

- Bayesian Network (BN): directed acyclic graph (DAG) $\mathcal{G}$ whose nodes are RVs and edges represent the influence of one node on another.
- Two viewpoints:
  - Data structure that is the skeleton for representing a joint probability distribution compactly in a factorized manner.
  - Compact representation for the conditional independence assumptions about a distribution.
- Two viewpoints are equivalent!

Student Example

- Student requests professor to write reference letter.



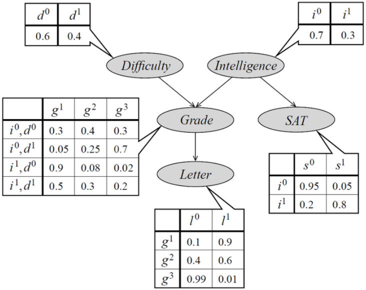
- Difficulty (of course): $Val(D) = \{d^0, d^1\} = \{easy, hard\}$.
- Intelligence (of student): $Val(I) = \{i^0, i^1\} = \{low, high\}$.
- SAT (scores of student): $Val(S) = \{s^0, s^1\} = \{low, high\}$.
- Grade (in course): $Val(G) = \{g^1, g^2, g^3\} = \{A, B, C\}$.
- Letter (quality): $Val(L) = \{l^0, l^1\} = \{weak, strong\}$.

Reasoning Patterns 1

- Causal reasoning or *prediction*: predicting downstream effects of various factors.

- $P_{\mathcal{B}student}(l^1 | i^0) = ?$
- $P_{\mathcal{B}student}(l^1 | i^0, d^0) = ?$

- Compute these values.

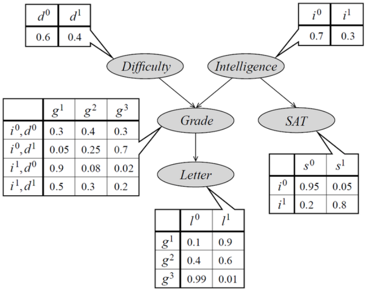


Reasoning Patterns 2

- Evidential reasoning or *explanation*: reason from effects to causes.

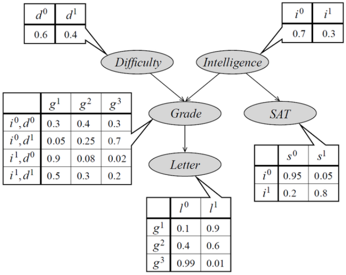
- $P_{\mathcal{B}^{student}}(i^1 | g^3) = ?$
- $P_{\mathcal{B}^{student}}(i^1 | l^0) = ?$
- $P_{\mathcal{B}^{student}}(i^1 | g^3, l^0) = ?$

- Compute these values.



Reasoning Patterns 3

- Intercausal reasoning or *explaining away*.
- $P_{\mathcal{B}student}(i^1|g^3) \neq P_{\mathcal{B}student}(i^1|g^3, d^1)$; why?
- Related to *v-structures*.



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	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

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

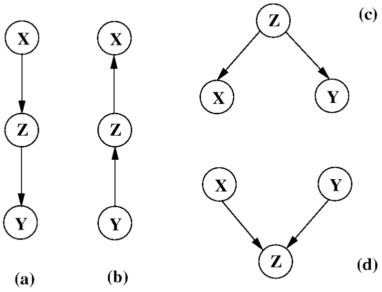
- Does D influence estimate of I through G ?

Local Independencies

- Bayesian network $\mathcal{G}$ is a DAG whose nodes represent RVs $X_1, \ldots, X_n$. Let $Pa^{\mathcal{G}}_{X_i}$ denote parents of X_i in $\mathcal{G}$.
- $\mathcal{G}$ encodes conditional independence assumptions, called local independencies $\mathcal{I}_l(\mathcal{G})$:
$$X_i : (X_i \perp NonDescendants_{X_i} | Pa^{\mathcal{G}}_{X_i})$$
- *Each node is conditionally independent of non-descendants given parents.*
- $P(I, D, G, S, L) = P(I)\, P(D)\, P(G|I, D)\, P(S|I)\, P(L|G)$.

Two-Edge Trails

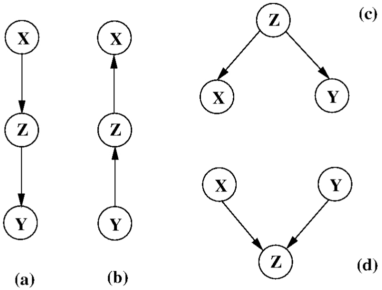
- Four possible two-edge trails from X to Y via Z.
- (a) Indirect causal effect; (b) indirect evidential effect.
- (c) A common cause; and (d) a common effect.



- When evidence can flow from X to Y via Z, the trail $X \rightleftharpoons Z \rightleftharpoons Y$ is active.

Active Trails

- (a) **Causal trail**
  $X \to Z \to Y$: active iff Z is not observed.
- (b) **Evidential trail**
  $X \leftarrow Z \leftarrow Y$: active iff Z is not observed.
- (c) **Common cause**
  $X \leftarrow Z \to Y$: active iff Z is not observed.



- (d) **Common effect** $X \to Z \leftarrow Y$: active iff Z or one of its descendants is observed.

D-Separation 1

- Let $\mathcal{G}$ be a BN, $X_1 \rightleftharpoons \ldots \rightleftharpoons X_n$ be a trail in $\mathcal{G}$ and **Z** be a subset of observed variables.
- Trail $X_1 \rightleftharpoons \ldots \rightleftharpoons X_n$ is active if:
  - For any v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, X_i or one of its descendant is in **Z**.
  - No other node on the trail is in **Z**.

D-Separation 2

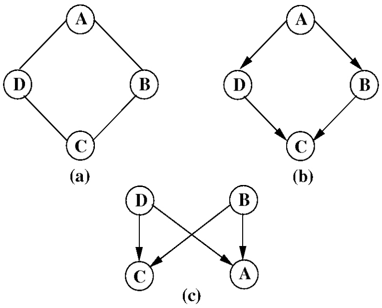
- Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ be three sets of nodes in $\mathcal{G}$. If there is no active trail between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$, $\mathbf{X}$ and $\mathbf{Y}$ are *d-separated* given $\mathbf{Z}$: $\text{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$.

- Global Markov independencies: set of independencies that correspond to d-separation (*directed separation*).

$$\mathcal{I}(\mathcal{G}) = \{(\mathbf{X} \perp \mathbf{Y}|\mathbf{Z}) : \text{d-sep}_{\mathcal{G}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z})\}$$

Misconception Example

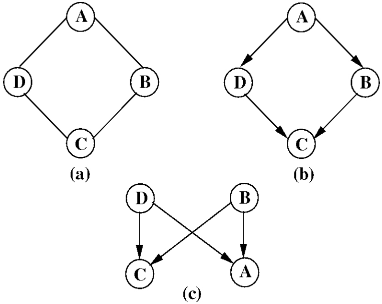
- Four students work in pairs on homework assignment.
- Only the following pairs meet: A and B; B and C; C and D; D and A—Figure (a).



- Students may have figured out professor's error from textbook. May transmit information to study partners.

Misconception Example

- For each $X \in \{A, B, C, D\}$, x^1/x^0 represent presence/absence of misconception.
- Need to model: $(A \perp C | B, D)$, $(B \perp D | A, C)$.
- Cannot represent with Bayesian network—Figures (b), (c).



- Can be modeled as a Markov network.

Coming up later...

- Bayes filters: Kalman filter, Particle filter.
- Localization and Mapping (SLAM).
- Decision making: MDPs (RL) and POMDPs.
- Please be prepared!