BAYES NETS

Presented by Zi Wang

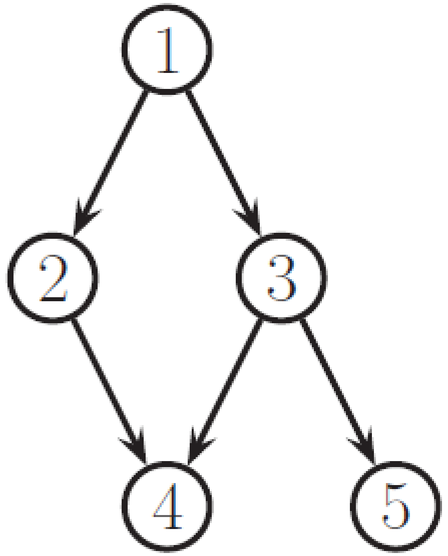
This talk covers

- *basic knowledge and must-know concepts*
- *some selected examples*
- *how to tell conditional independence*
- *exercises on independence quiz*
- *decision diagrams*

A review of MLPP chapter 10

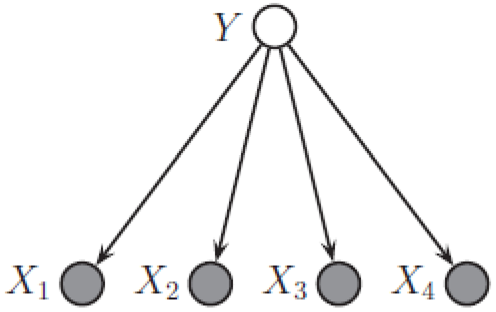
BASIC KNOWLEDGE

- Chain rule: $p(x_{1:V}) = p(x_1)p(x_2|x_1)p(x_3|x_2,x_1)\ p(x_4|x_3,x_2,x_1)\ldots p(x_V|x_{1:V})$
- Conditional independence: $X \perp \mathrm{Y}|\mathrm{Z} \Leftrightarrow \mathrm{p}(\mathrm{X},\mathrm{Y}|\mathrm{Z}) = \mathrm{p}(\mathrm{X}|\mathrm{Z})$
- Graphical models:
  - Node (vertice): random variables
  - Link (edge)
  - Parent, parents + itself = family
  - Child
  - Root
  - Leaf
  - Ancestor
  - Descendent
  - DAG: directed acyclic graph
  - Clique
  - …

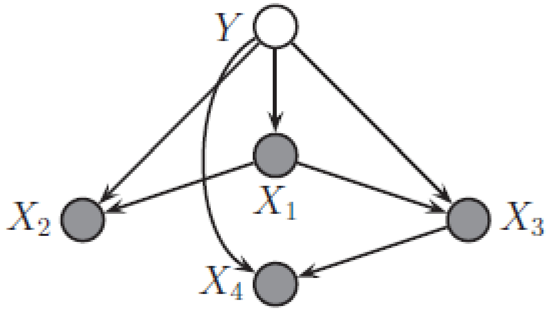


SELECTED EXAMPLES

- Naïve Bayes classifiers

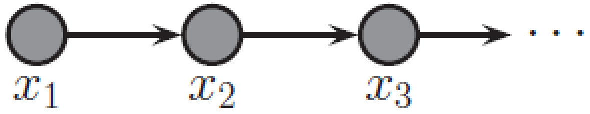


capture features' correlation



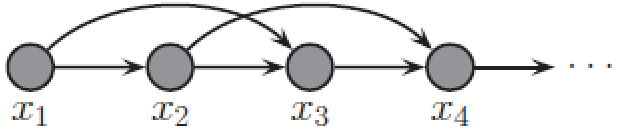
Generalize to DGM

- Markov chain
  - First order



$$p(x_{1:V}) = p(x_1)\prod_{t=1}^{V} p(x_t|x_{t-1})$$

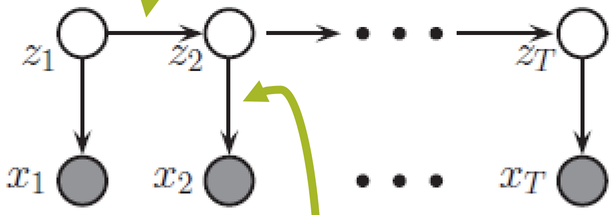
  - Second order



$$p(x_{1:V}) = p(x_1, x_2)\prod_{t=1}^{V} p(x_t|x_{t-1}, x_{t-2})$$

- HMM

Transition model



observation model

MEDICAL DIAGNOSIS

- Alarm network
  - Question: given DGM's structure and data, how to learn parameters?
- Quick medical reference
  - Hidden nodes: disease (prior); visible nodes: symptoms
  - Two ways to model p(symptom | the symptom's parent diseases)
    - logistic regression: $p(v_t | h_{pa(t)}) = \frac{1}{1+e^{-w_t^T h_{pa(t)}}}$
    - Noisy-OR model, including a dummy leak node $h_0 \equiv 1, \forall t, \mathrm{h}_0 \in pa(t)$
      - $q_{st} = p(v_t = 0 | h_s = 1, h_{-s} = 0)$, $\theta_{st} = 1 - q_{st}$
      - => $p(v_t = 0 | \boldsymbol{h}) = \theta_{0t} \prod_{s \in pa(t)} \theta_{st}^{h_s}$ => $p(v_t = 1 | \boldsymbol{h}) = 1 - p(v_t = 0 | \boldsymbol{h}) = 1 - \mathrm{e}^{w_t^T h_{pa(t)}}$

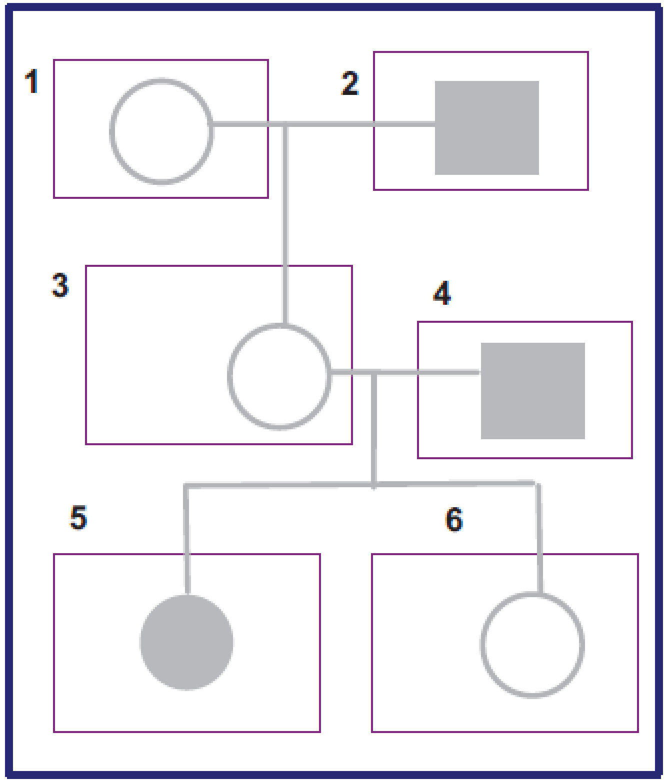
GENETIC LINKAGE ANALYSIS

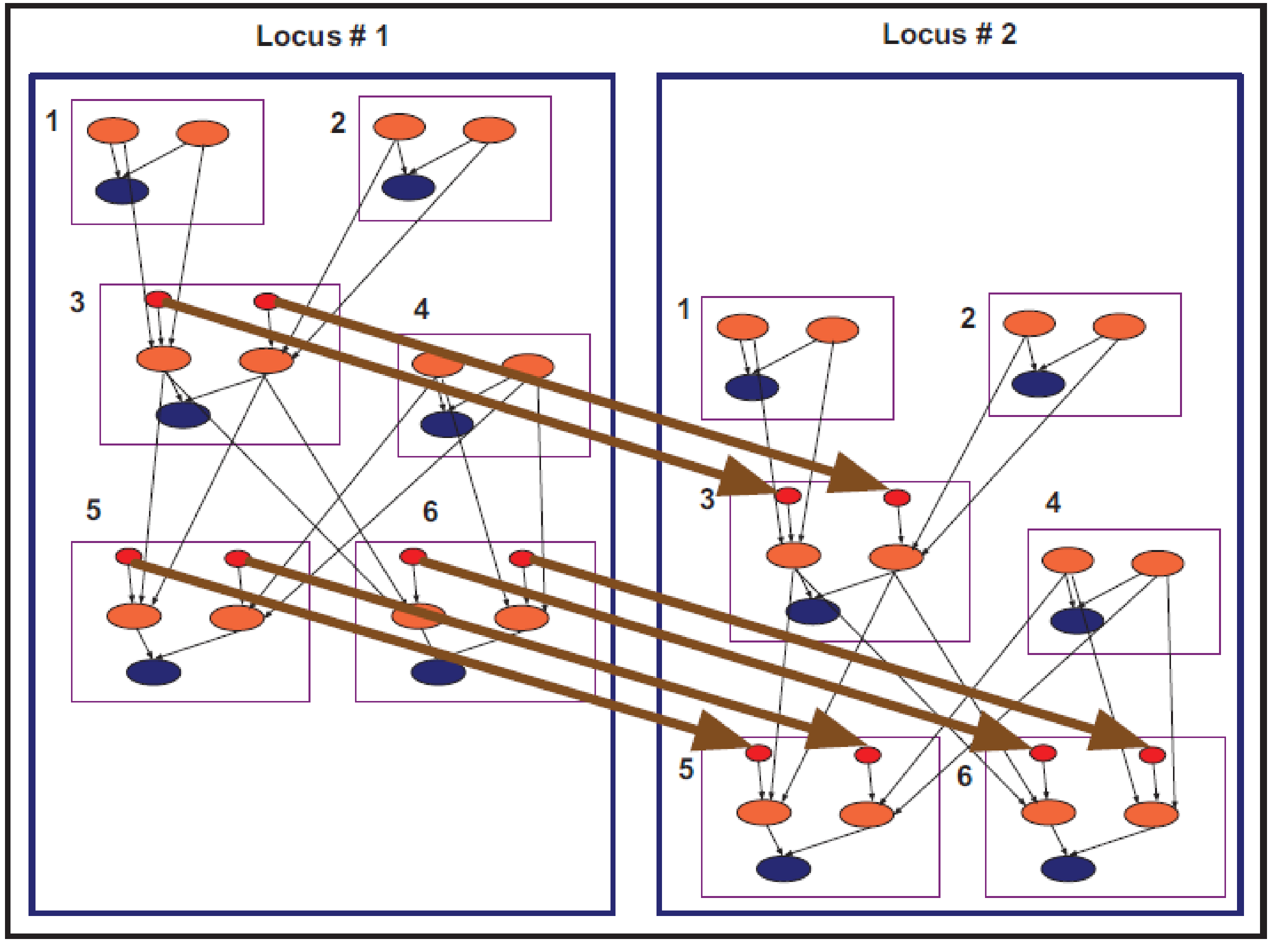
Terms:

- Marker (like blood type)
- Genotype (maternal allele, paternal allele)
- Inheritance model (Mendelian)

$$p(G_{ij}^m | G_{kj}^m, G_{kj}^p, Z_{ij}^m) = \begin{cases} \mathbb{I}(G_{ij}^m = G_{kj}^m) & \text{if } Z_{ij}^m = m \\ \mathbb{I}(G_{ij}^m = G_{kj}^p) & \text{if } Z_{ij}^m = p \end{cases}$$

GENETIC LINKAGE ANALYSIS





INFERENCE

- Problem definition:
  - $x_{1:V}$ is the random variables with $\boldsymbol{x_v}$ visible and $\boldsymbol{x_h}$ hidden
  - **Given parameter $\boldsymbol{\theta}$** , the joint distribution $\mathrm{p}(\mathbf{x}|\boldsymbol{\theta})$
  - Infer the posterior distributions of the unknown from the known

$$p(x_h|x_v,\theta) = \frac{p(x_h,x_v|\theta)}{p(x_v|\theta)}, p(x_v|\theta) = \sum_{x_h} p(x_h,x_v|\theta)$$

  - Infer the posterior distributions of what we ‘re interested in but the unknown

$$p\left(x_q \middle| x_v,\theta\right) = \sum_{x_n} p(x_q,x_n|x_v,\theta)$$

LEARNING

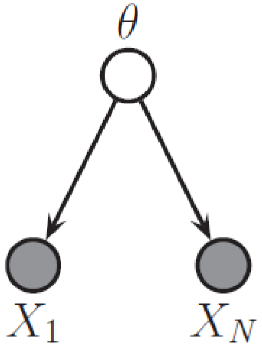
- Problem: the parameters $\boldsymbol{\theta}$ is unknown and we need to learn $\boldsymbol{\theta}$
- Solution: MAP

$$p(\theta|\boldsymbol{x}) \propto p(\boldsymbol{x}|\theta)p(\theta)$$

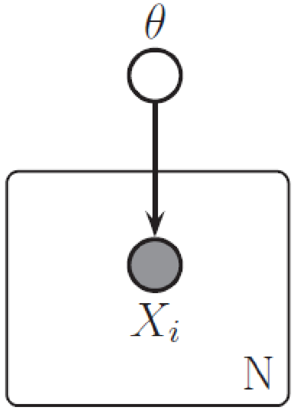
$$\boldsymbol{\theta} = arg\max_{\theta} \log(p(\theta|\boldsymbol{x}) = arg\max_{\theta} \log p(\boldsymbol{x}|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta})$$

- Add parameters as nodes
  - Hidden variables vs. parameter

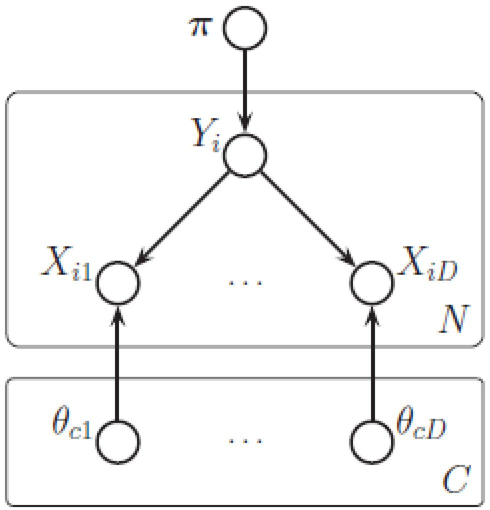
PLATE



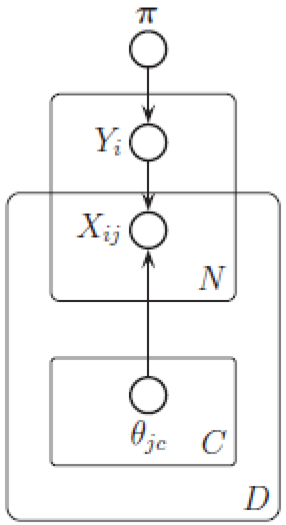
θ
X_1
X_N



θ
X_i
N



π
Y_i
X_{i1}
$\dots$
X_{iD}
N
θ_{c1}
$\dots$
θ_{cD}
C



π
Y_i
X_{ij}
N
θ_{jc}
C
D

LEARNING FROM COMPLETE DATA

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} p(\mathbf{x}_i|\boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{t=1}^{V} p(x_{it}|\mathbf{x}_{i,\mathrm{pa}(t)}, \boldsymbol{\theta}_t) = \prod_{t=1}^{V} p(\mathcal{D}_t|\boldsymbol{\theta}_t)$$

$$p(\boldsymbol{\theta}) = \prod_{t=1}^{V} p(\boldsymbol{\theta}_t)$$

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = \prod_{t=1}^{V} p(\mathcal{D}_t|\boldsymbol{\theta}_t)p(\boldsymbol{\theta}_t)$$

D-SEPARATION'S MOTIVATION

- Blood sugar --> stomach acidity --> hunger
- What are the testable statistical consequences of causal structure?
- 1930~1960: Wright, Blalock and Simon considered vanishing correlations and vanishing partial correlations => conditional independence
- What do we want?
- **Compute the independence and conditional independence relations that hold for all causal strengths!**

http://www.andrew.cmu.edu/user/scheines/tutor/d-sep.html
http://www.aispace.org/downloads.shtml

D-SEPARATION INTUITION

- D – dependence
- Active path
- 1) A --> C --> B
  2) A <-- C <-- B
  3) A <-- C --> B
  4) A --> C <-- B

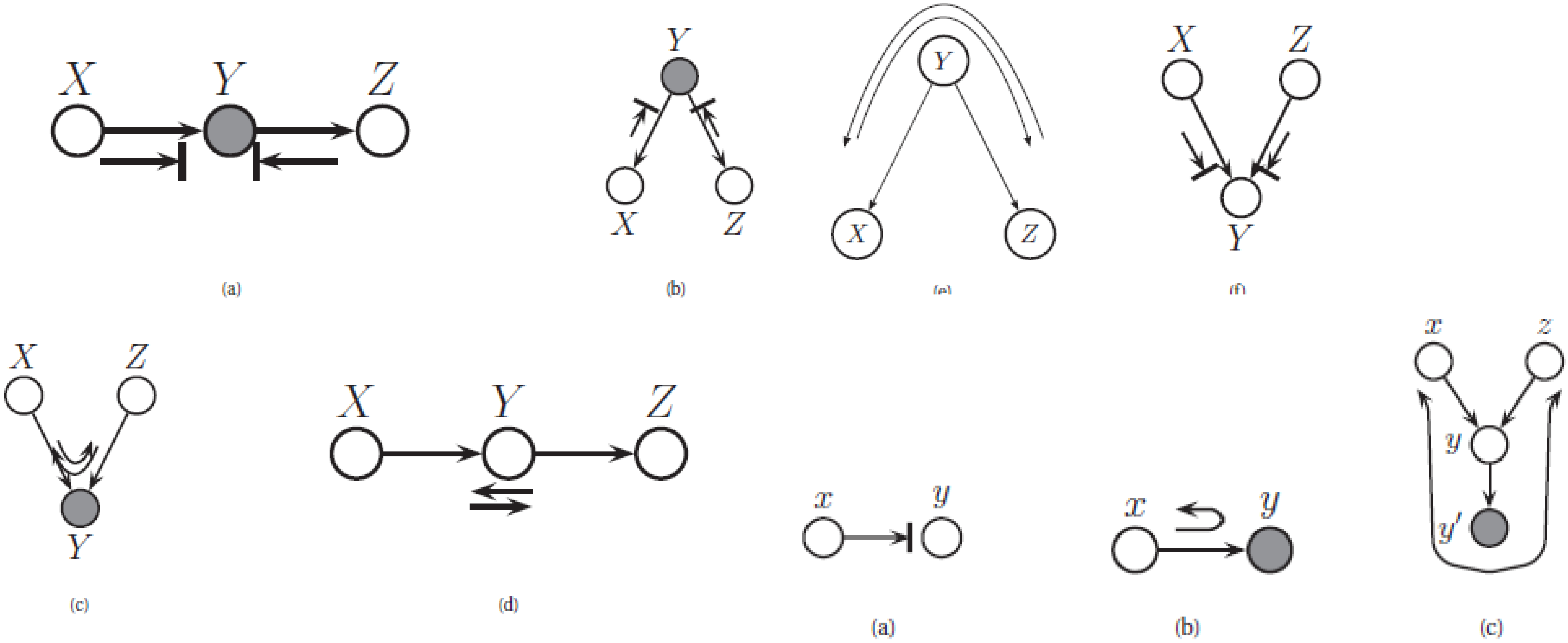
Active path, C is non-collider

Inactive path, C is collider

- dead battery --> **car won't start** <-- no gas (by Pearl, 1988)

http://www.andrew.cmu.edu/user/scheines/tutor/d-sep.html
http://www.aispace.org/downloads.shtml
http://bayes.cs.ucla.edu/BOOK-2K/d-sep.html

BAYES BALL RULE



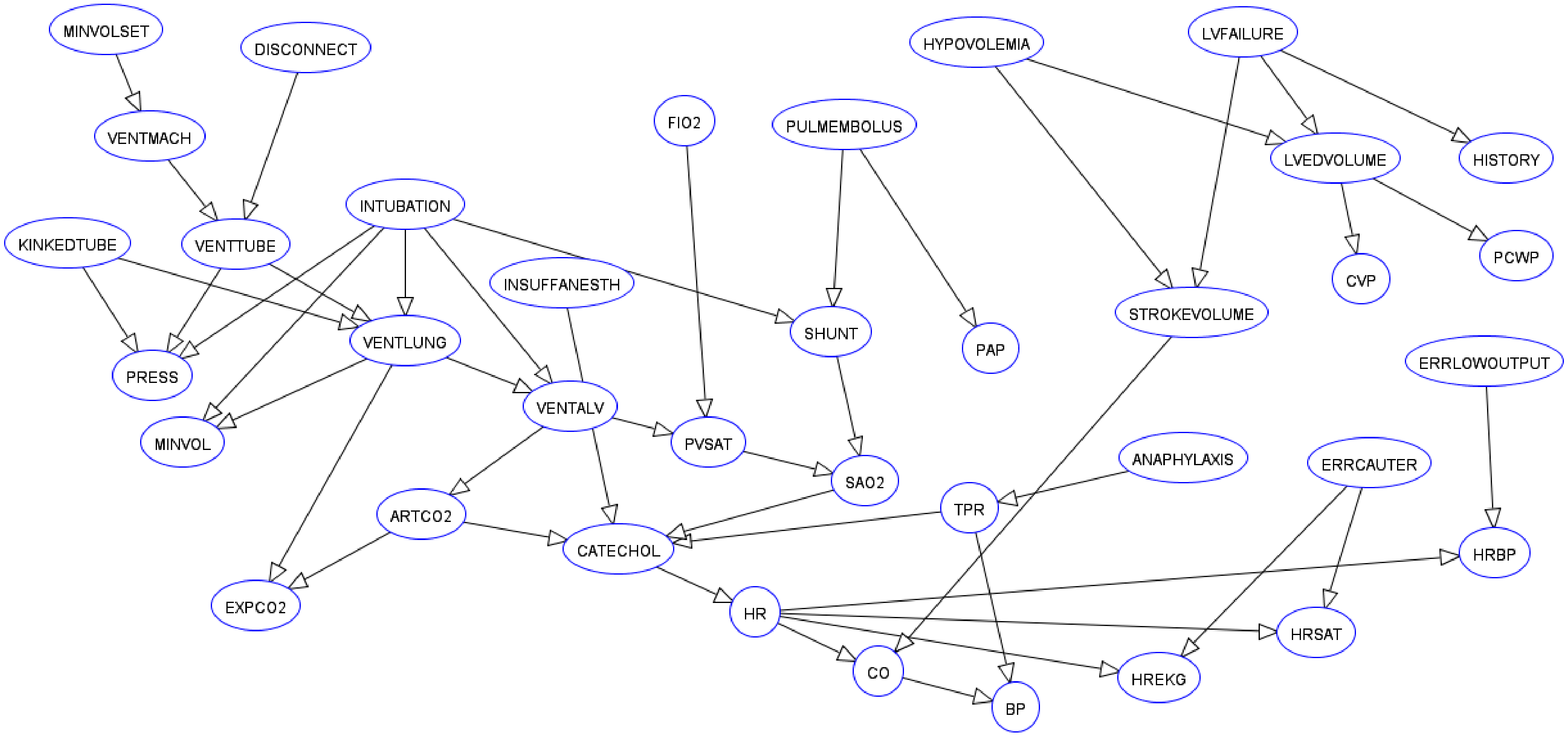
D-SEPARATION DEFINATION

- **D-connection**:
  If G is a directed graph in which X, Y and Z are disjoint sets of vertices, then X and Y are d-connected by Z in G if and only if there **exists an undirected path U** between some vertex in X and some vertex in Y such that **for every collider C on U**, <u>either C or a descendent of C is in Z</u>, **and no non-collider on U is in Z**.

  X and Y are **d-separated** by Z in G if and only if they are not d-connected by Z in G.

http://www.andrew.cmu.edu/user/scheines/tutor/d-sep.html
http://www.aispace.org/downloads.shtml

EXERCISE TIME!



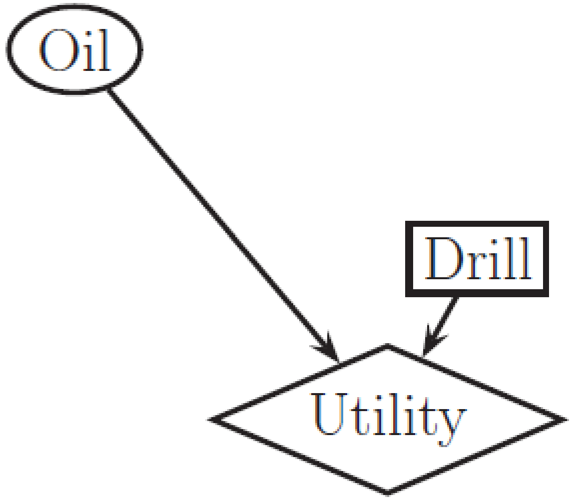
http://www.aispace.org/downloads.shtml

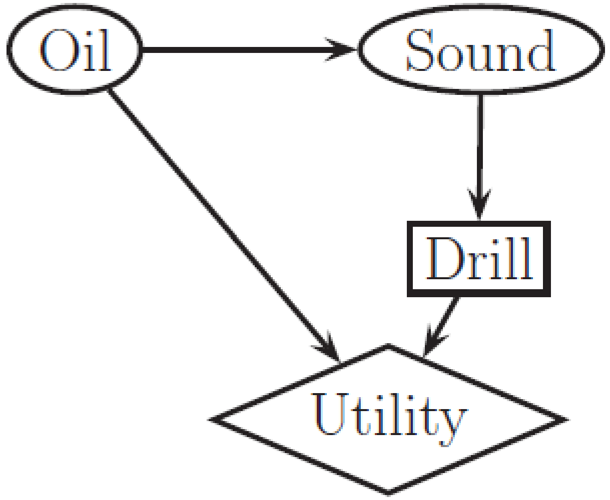
OTHER MARKOV PROPERTIES FOR DGM

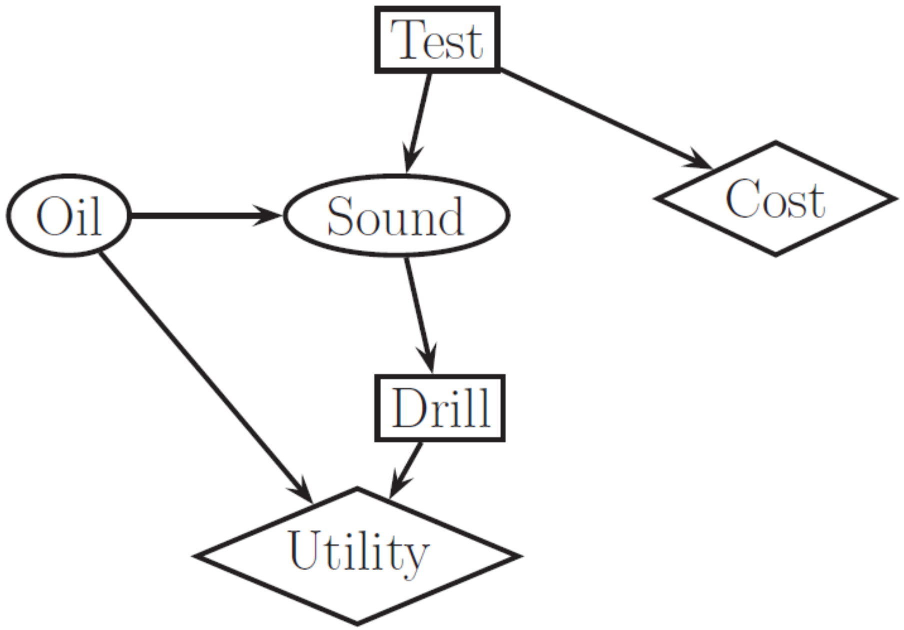
- Directed local markov property
  - $t \perp \mathrm{nd(t)\backslash pa(t)|pa(t)}$
- Ordered Markov property
  - $t \perp \mathrm{pred(t)\backslash pa(t)|pa(t)}$
- Markov blanket: the set of nodes consisting of its parents, its children, and any other parents of its children

http://en.wikipedia.org/wiki/Bayesian_network#Markov_blanket

DECISION DIAGRAM







SUMMARY

- Basic ideas about bayes nets
- Basic ideas about inference and learning for bayes nets
- **How to tell whether X and Y are conditionally independent given Z in G**
- Influence diagram