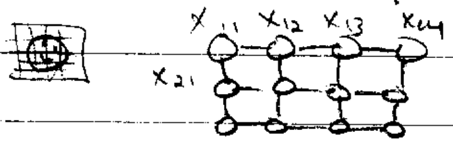
- Assignment: +1 week, covars plots, notes on web
- Review Bayes Ball, d-sep
- Want efficient algos on graph

Undirected Graphs

Sometimes can't consider parent-child types of links.

i.e. image pixels: "MRFs"

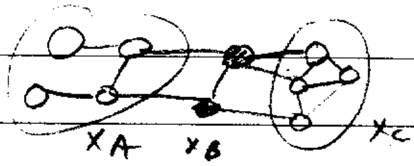


0 = dark
1 = bright

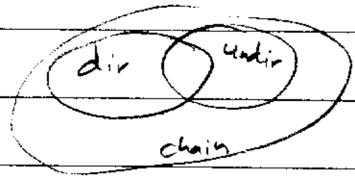
pixels nearby are dependent because images have smooth regions of similar gray levels.

again want $p(x_1, \dots, x_n)$ pdf...

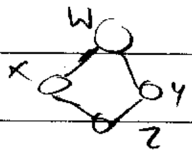
$X_A \perp\!\!\!\perp X_C \mid X_B$ is true if paths from X_A to X_C all cross X_B. (standard graph theory separation)



Undirected $\not\subseteq$ Directed   Directed $\not\subseteq$ Undirected



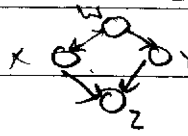
example



$\equiv$ $X \perp\!\!\!\perp Y \mid \{W, Z\}$
$W \perp\!\!\!\perp Z \mid \{X, Y\}$

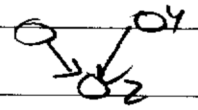
but in directed can't

in directed, acyclic, at least one node will be "V" let Z be that node



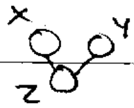
$\equiv$ $X \perp\!\!\!\perp Y \mid W$
$\cancel{X \perp\!\!\!\perp Y \mid \{W, Z\}}$ since Ball goes through

example



$\equiv$ $X \perp\!\!\!\perp Y$
$X \not\perp\!\!\!\perp Y \mid Z$

for undirected...



$X \perp\!\!\!\perp Y \mid Z$

$X \perp\!\!\!\perp Y$

undirected $p(x_1,\dots,x_n) \propto \prod_{c\in\mathcal{C}} \Psi_{x_c}(x_c)$

$\Leftarrow$ +ve functions | potentials

else could have $\prod_{c\in\mathcal{C}} \Psi_{x_c}(x_c)$ $\prod_{c\in\mathcal{C}} \Psi_{x_c}(x_c)$

max cliques, clique

redundant

$C \equiv$ maximal clique, fully interconnected subset of nodes.

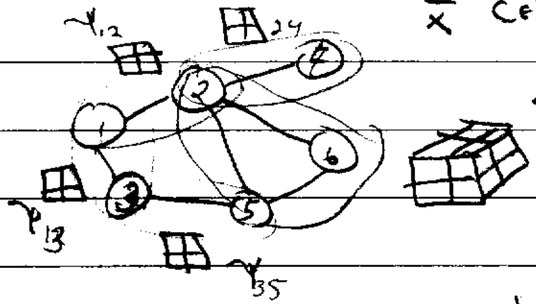
non maximal

maximal cliques: no member is a subset of another member.

Unlike directed $p(x_i \mid x_{\pi_i})$, $\sum_{x_c} \Psi_{x_c}(x_c) \neq 1$

$p(x_1,\dots,x_n) = \frac{1}{Z} \prod_{c\in\mathcal{C}} \Psi_{x_c}(x_c)$

$Z = \sum_{x} \prod_{c\in\mathcal{C}} \Psi_{x_c}(x_c)$ since $\sum_{x_v} p(x_v) = 1$



Z = normalization const. or partition function (hard to compute)

potentials need not be marginals | conditionals

Since $\Psi_{x_c}(x_c)$ +ve, $\Psi_{x_c}(x_c) = \exp(-H_{x_c}(x_c))$

energy

Statistical physics — spins $x_i \in \{\pm 1\}$ — spin up spin down, magnetic behavior of crystals

Ising model

$\hookrightarrow$ nobel prize: analytic for multi dims (3D+, impossible?)

800 papers per year published

$\Leftarrow$ Boltzmann distrib $p(x) = \frac{1}{Z}\exp(-H(x))$

exact solution to distrib 1D (Ising 1926), 2D (Onsager, 1944), 3D?

homogeneous toroidal

$p(x) = \frac{1}{Z}\exp(-T\, H(x))$

$\Leftarrow$ temperature

- So far, have efficient way to check $X_A \perp\!\!\!\perp X_C \mid X_B$? for Both directed & undirected

- Want efficient way to compute $p(x_F | x_E)$

  where X is split into $\{X_E, X_F, X_H\}$

  un-shading of X_F (queried conditionals)

  light shading of X_H (marginalized)

  dark shading of X_E (observed)

$$p(x_F | x_E) = \frac{\sum_{x_H} p(x_E, x_F, x_H)}{\sum_{x_F} \sum_{x_H} p(x_E, x_F, x_H)} \quad \begin{array}{l} \leftarrow \text{sum over } c^{|H|} \\ \leftarrow \text{sum over } c^{|H|+|F|} \end{array}$$

i.e. $p(x_1, \ldots, x_5) = \sum_{x_6} p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) p(x_5|x_3) p(x_6|x_2, x_5)$

$= p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) p(x_5|x_3) \sum_{x_6} p(x_6|x_2, x_5)$

just unity but lets say we do it

$O(c^6)$

$O(c^3)$ entries to sum over

x_1

0.3	0.7

want $p(x_1 | \bar{x}_6)$ for a specific value of x_6 is 1 dimension slice of pdf

$$p(x_1, \bar{x}_6) = \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} p(x_1) p(x_2|x_1) p(x_3|x_1) p(x_4|x_2) p(x_5|x_3) p(\bar{x}_6 | x_2, x_5)$$

$O(c^3)$ $= p(x_1) \sum_{x_2} p(x_2|x_1) \sum_{x_3} p(x_3|x_1) \sum_{x_4} p(x_4|x_2) \underbrace{\sum_{x_5} p(x_5|x_3) p(\bar{x}_6 | x_2, x_5)}_{\phi_{x_5}(x_2, x_3)}$

no other x_4 depends

$O(c^2)$ $= p(x_1) \sum_{x_2} p(x_2|x_1) \sum_{x_3} p(x_3|x_1) \; \phi_{x_4}(x_2) \; \phi_{x_5}(x_2, x_3)$

$O(c^3)$ $= p(x_1) \sum_{x_2} p(x_2|x_1) \; \phi_{x_4}(x_2) \; \phi_{x_3}(x_1, x_2)$ just unity but lets say we do it

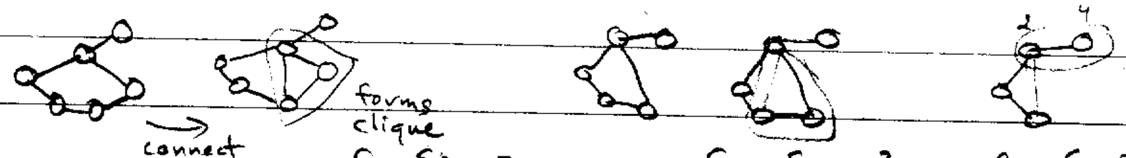
$O(c^2)$ $= p(x_1) \; \phi_{x_2}(x_1)$

$$p(\bar{x}_6) = \sum_{x_1} p(x_1) \phi_{x_2}(x_1) \quad \leftarrow O(c^1)$$

$$p(x_1 | \bar{x}_6) = \frac{p(x_1) \phi_{x_2}(x_1)}{\sum_{x_1} p(x_1) \phi_{x_2}(x_1)}$$

How to automate this?
~~first undirected algo.~~
then directed algo.

- Undirected Elimination
  - pick an elimination ordering (assume given, all are valid but can have variable efficiency)
    eg. $X_6, X_5, X_4, X_3, X_2, X_1$
  - for each X_i
    connect all neighbours of X_i
    eliminate X_i

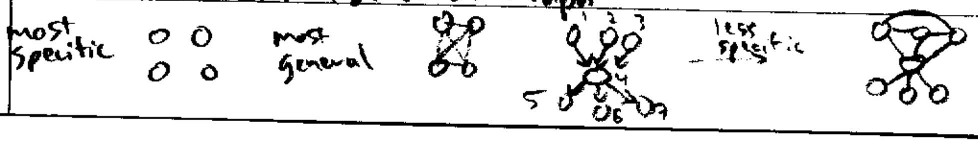


$C_6 = \{2,5,6\}$ $\quad C_5 = \{2,3,5\}$ $\quad C_4 = \{2,4\}$

$X_{C_6} = \{X_2, X_5, X_6\}$ $\quad X_{C_5} = \{X_2, X_3, X_5\}$

- Largest clique size determines complexity, $O(c^k)$
  $\triangleq$ Tree width of graph $\equiv k$
- marginalizing a variable, sums over prod of all factors that use it
  couples all the factors (i.e. the neighbours)
  $\sum_{x_5} p(x_5 | x_3)\, p(x_6 | x_2, x_5)$
  induces dependency between X_2 & X_3
  (also, irrelevant if X_i is to left or right of bar $p(\cdot|\cdot)$, dependency coupling is undirected)

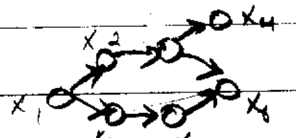
- Directed Graph Elimination
  can make a directed graph undirected (but MORE general)
  by marrying parents (moralization) & dropping arrow heads
  then run the same algo. as above
  Moralization & Moral Graph



$\psi(1,2,3,4)\,\psi(4,5)\,\psi(5,6)\,\psi(4,7)$

more general graph contains the previous one (but fewer $\perp\!\!\!\perp$ statements)

i.e. 

e.g. if in 

x_5 is eliminated with moralizing then X_2 is not its neighbour and that clique is lost.

X_F single node

More general Junction Tree Algorithm

- General Directed Elimination: $p(X_F|\bar{X}_E)$ — slice pdf, evidential function

$$f(\bar{x}_i) = \sum_{x_i} f(x_i)\,\delta(x_i,\bar{x}_i)$$

$$\phi_{x_6}(x_2,x_5) = \sum_{x_6} p(x_6|x_2,x_5)\,\delta(x_6,\bar{x}_6) = p(\bar{x}_6|x_2,x_5)$$

potential functions active list = $\begin{cases} \text{all evidential functions } \delta(x_i,\bar{x}_i) \\ p(x_i|x_{\pi_i}) \\ \text{intermediate } \phi_{x_6}(x_2,x_5) \end{cases}$

1) start with list = $\begin{cases} \text{evidentials} \\ p(x_i|x_{\pi_i}) \end{cases}$

2) choose an ordering of nodes where query appears last

3) for each i:
   a) find all potentials containing x_i & remove from list
   b) ψ_{x_i} = product of all pots.
   c) $\phi_{x_i} = \sum_{x_i} \psi_{x_i}$
   d) put ϕ_{x_i} on active list

terminate when we get to query node X_F

get $p(X_F,\bar{X}_E) = \psi_{X_F}(X_F)$

$$p(X_F|\bar{X}_E) = \frac{\psi_{X_F}(X_F)}{\phi_{X_F}}$$

$$\phi_{x_i}(X_{S_i}) = \sum_{x_i} \psi_{x_i}(X_{C_i}) \qquad X_{S_i} \triangleq X_{C_i} \setminus X_i$$

- Undirected: same as above yet $p(x_i|x_{\pi_i})$ replaced with $\psi_{X_C}(X_C)$ potentials. Also normalization term Z is kept around but cancels out for conditionals (but not marginals) $p(X_F|\bar{X}_E) = \dfrac{\frac{1}{Z}\psi_{X_F}(X_F)}{\frac{1}{Z}\sum_{X_F}\psi_{X_F}(X_F)}$

- To compute Z, $p(X_F,\bar{X}_E) = \dfrac{\psi_{X_F}(X_F)}{Z}$, normalize to 1