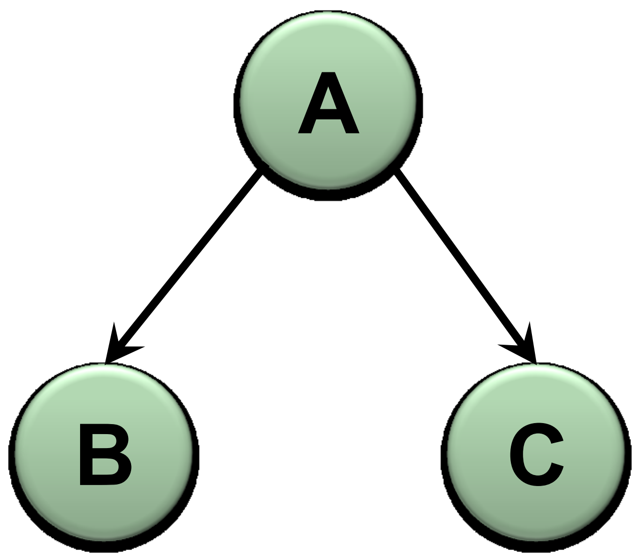
Bayesian Networks Recitation

Anthony Gitter

November 8, 2010

Bayesian networks basics

- Directed acyclic graphs
- Edges represent dependencies
- Exploit independence to compactly represent join distribution



Important property of probability

- $X \perp Y \mid Z$ means $P(X,Y|Z) = P(X|Z)P(Y|Z)$

- If $X \perp Y \mid Z$ then $P(X|Y,Z) = P(X|Z)$
  - Why?

- $$\begin{aligned} P(X,Y|Z) &= P(X,Y,Z) / P(Z) \\ &= P(X|Y,Z)P(Y,Z) / P(Z) \\ &= P(X|Y,Z)P(Y|Z) \end{aligned}$$

- $P(X,Y|Z) = P(X|Y,Z)P(Y|Z)$
- $P(X,Y|Z) = P(X|Z)P(Y|Z)$ when $X \perp Y \mid Z$
- This implies $P(X|Z) = P(X|Y,Z)$ when $X \perp Y \mid Z$

How does conditional independence help?

- Can always use chain rule to factor a distribution

- Earthquake example:
- P(A,B,E,J,M) =

  P(M|A,B,E,J)P(A,B,E,J) =

  P(M|A,B,E,J)P(J|A,B,E)P(A,B,E) =

  P(M|A,B,E,J)P(J|A,B,E)P(A|B,E)P(B,E) =

  P(M|A,B,E,J)P(J|A,B,E)P(A|B,E)P(B|E)P(E)

- What Bayes net does this correspond to?

How does conditional independence help?

- Apply independence assumptions
  - Obtained through prior domain knowledge
  - Learned from the data

- A few earthquake independence assumptions

  $J \perp B,E \mid A$

  $M \perp B,E,J \mid A$

  $B \perp E$

- May be others as well that we won't need here

How does conditional independence help?

- $P(A,B,E,J,M) =$
  $P(M|A,B,E,J)P(J|A,B,E)P(A|B,E)P(B|E)P(E)$

$J \perp B,E \mid A \quad \rightarrow \quad P(J|A,B,E) = P(J|A)$

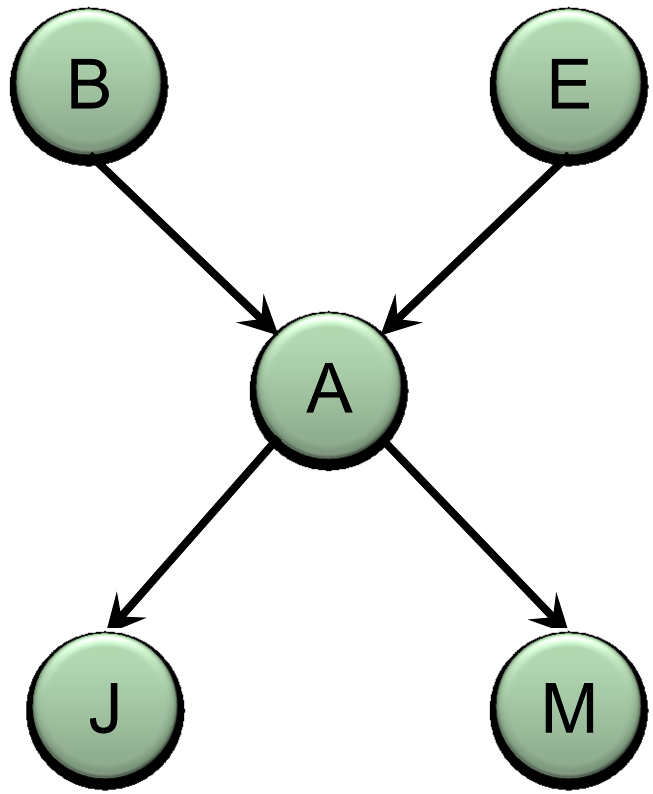
$M \perp B,E,J \mid A \quad \rightarrow \quad P(M|A,B,E,J) = P(M|A)$

$B \perp E \quad \rightarrow \quad P(B|E) = P(B)$

- $P(A,B,E,J,M) = P(M|A)P(J|A)P(A|B,E)P(B)P(E)$
  - Does this look familiar?

How does conditional independence help?

- $P(A,B,E,J,M) = P(M|A)P(J|A)P(A|B,E)P(B)P(E)$



- This factorization of the joint distribution corresponds to the Bayesian network from lecture

Chain rule order

- What if we apply the chain rule in a different order?

- P(A,B,E,J,M) =

  P(A | B,E,J,M) P(B,E,J,M) =

  P(A | B,E,J,M) P(B | E,J,M) P(E,J,M) =

  P(A | B,E,J,M) P(B | E, J,M) P(E | J,M) P(J,M)

  P(A | B,E,J,M) P(B | E, J,M) P(E | J,M)P(J | M)P(M)

- Get a different Bayes net structure even if we simplify further using independence assumptions
- General Bayes net structure learning is outside our scope

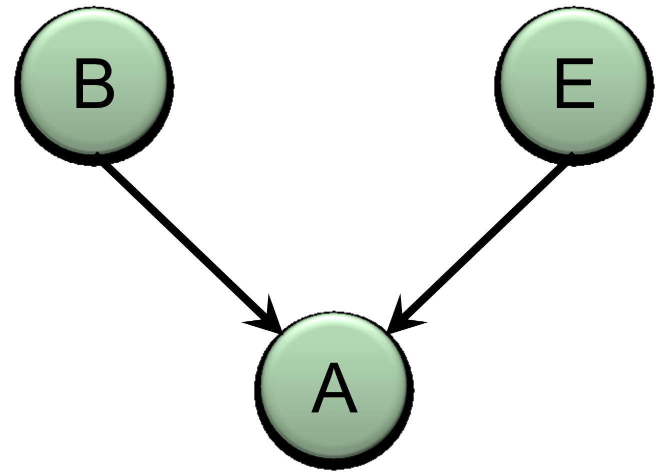
Bayes net structure and independence

- Edges that are not in the graph correspond to independence assumptions
  - Missing edges often most interesting
  - Complete graph can represent any distribution
  - Graph without edges makes very strong assumptions

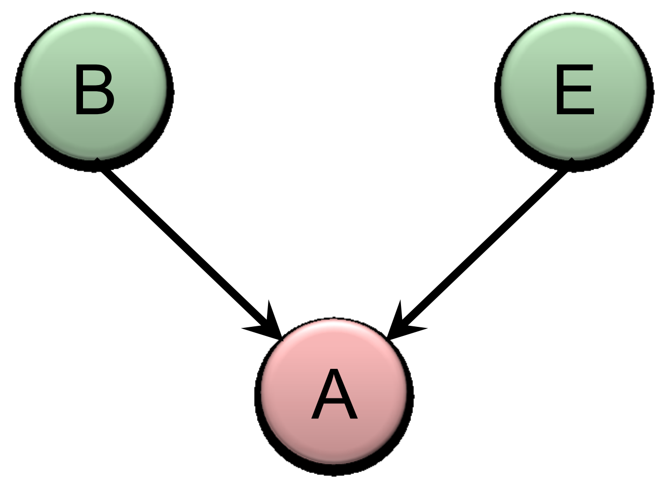
- We've seen independence assumptions → structure
- What about structure → independence assumptions?

Structure encodes independence

- $B \perp E$
- If there is not a burglary that doesn't tell us anything about whether there will be an earthquake



- However $B \not\perp E \mid A$
- If the alarm is on and there is no earthquake, there is likely a burglary

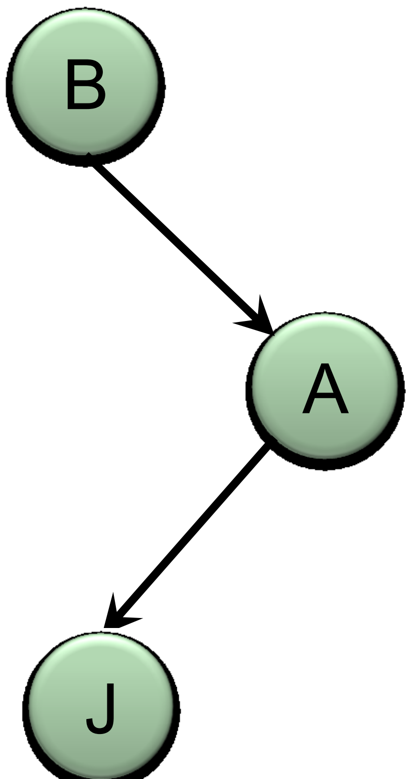


- This is called a collider or v-structure

Structure encodes independence

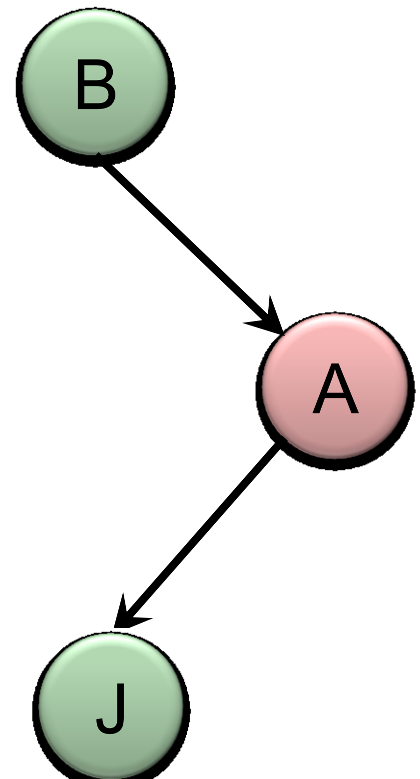
$B \not\perp J$

If there is a burglary it is more likely John will call



$B \perp J \mid A$

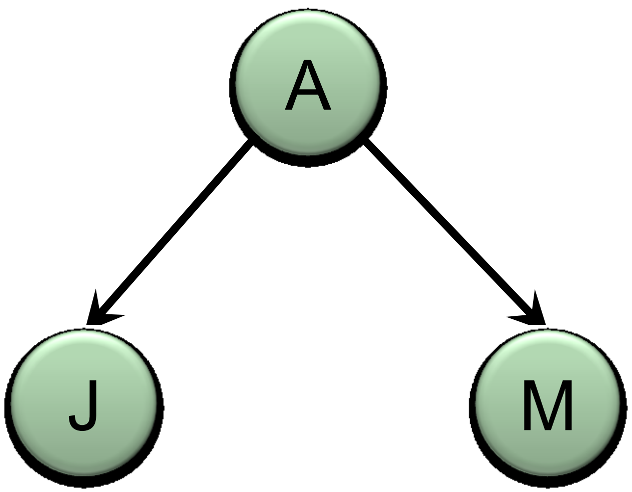
If we know the alarm is on burglary does not provide additional information



Structure encodes independence

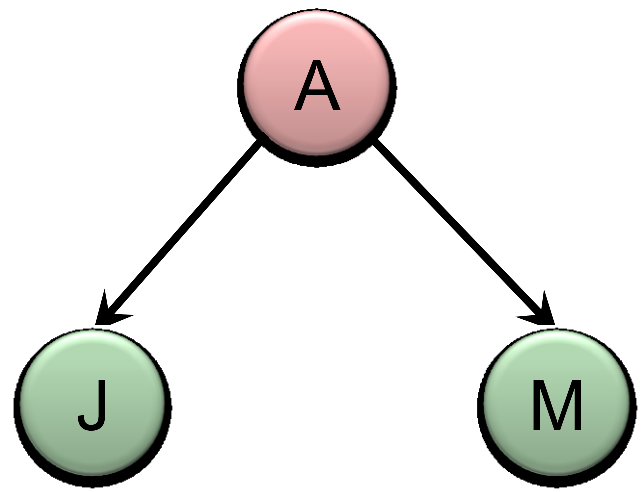
$J \not\perp M$

If John calls it is more likely that Mary will call



$J \perp M \mid A$

If the alarm is off, Mary calling is not influenced by John calling



d-separation

- d-separation formalizes this intution
- If X and Y are d-separated given Z, then $X \perp Y \mid Z$
- Z is a set of nodes
- To determine if X and Y are d-separated examine all undirected paths between them
  - Can use the directed edges in either direction
- If all paths are d-separated by Z, then X and Y are d-separated by Z

Some of this material from Wikipedia

d-separation

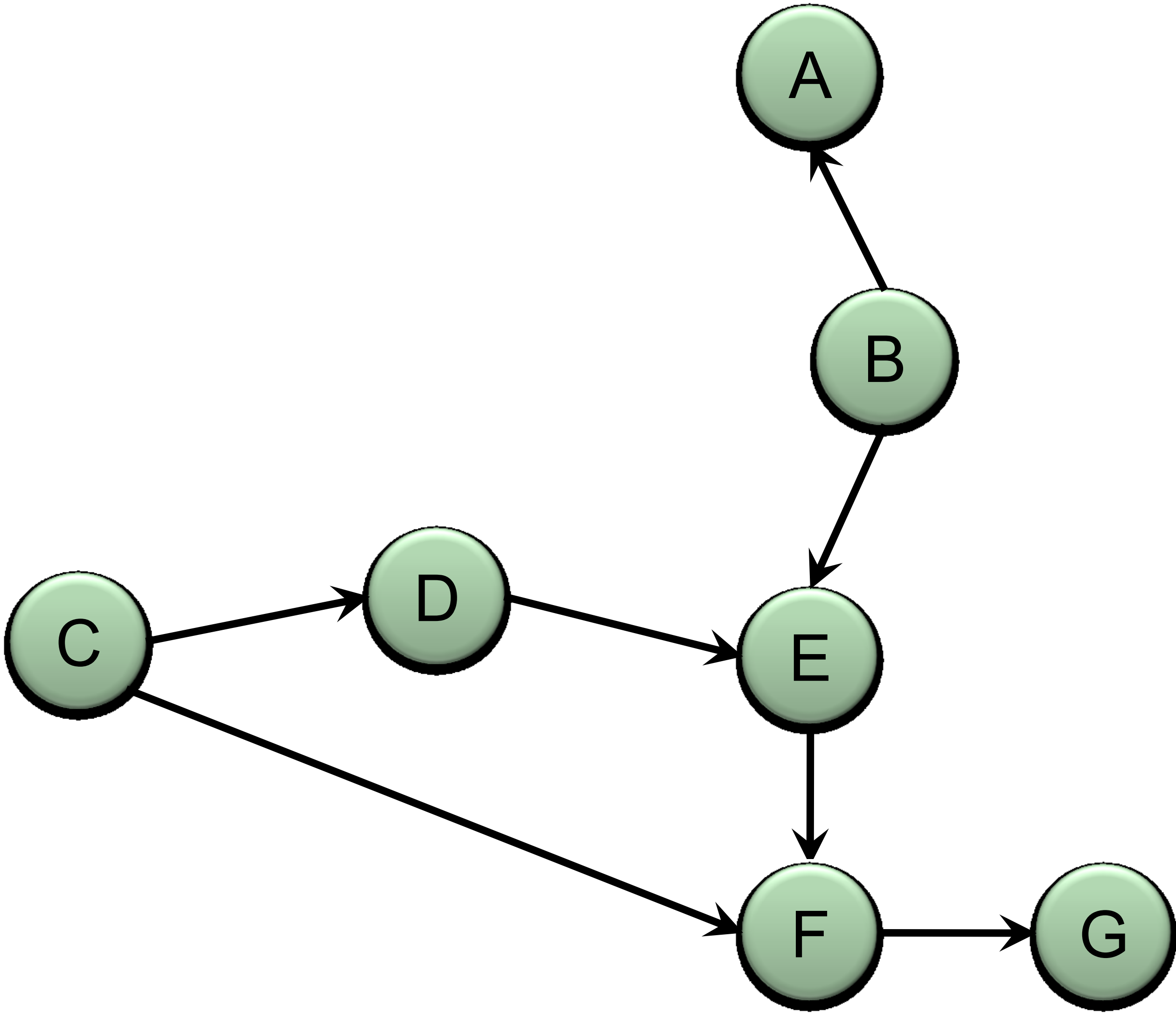
- A path between X and Y is d-separated given Z if at least one of the following is true:
  - The path contains $i \rightarrow j \rightarrow k$ and j is in Z
  - The path contains $i \leftarrow j \rightarrow k$ and j is in Z
  - The path contains $i \leftarrow j \leftarrow k$ and j is in Z
  - The path contains $i \rightarrow j \leftarrow k$ and j is *not* in Z and no descendant of j is in Z

- Path that does not d-separate X and Y is called *active*
- If there exists an active path, X and Y are d-connected given Z

Some of this material from Wikipedia

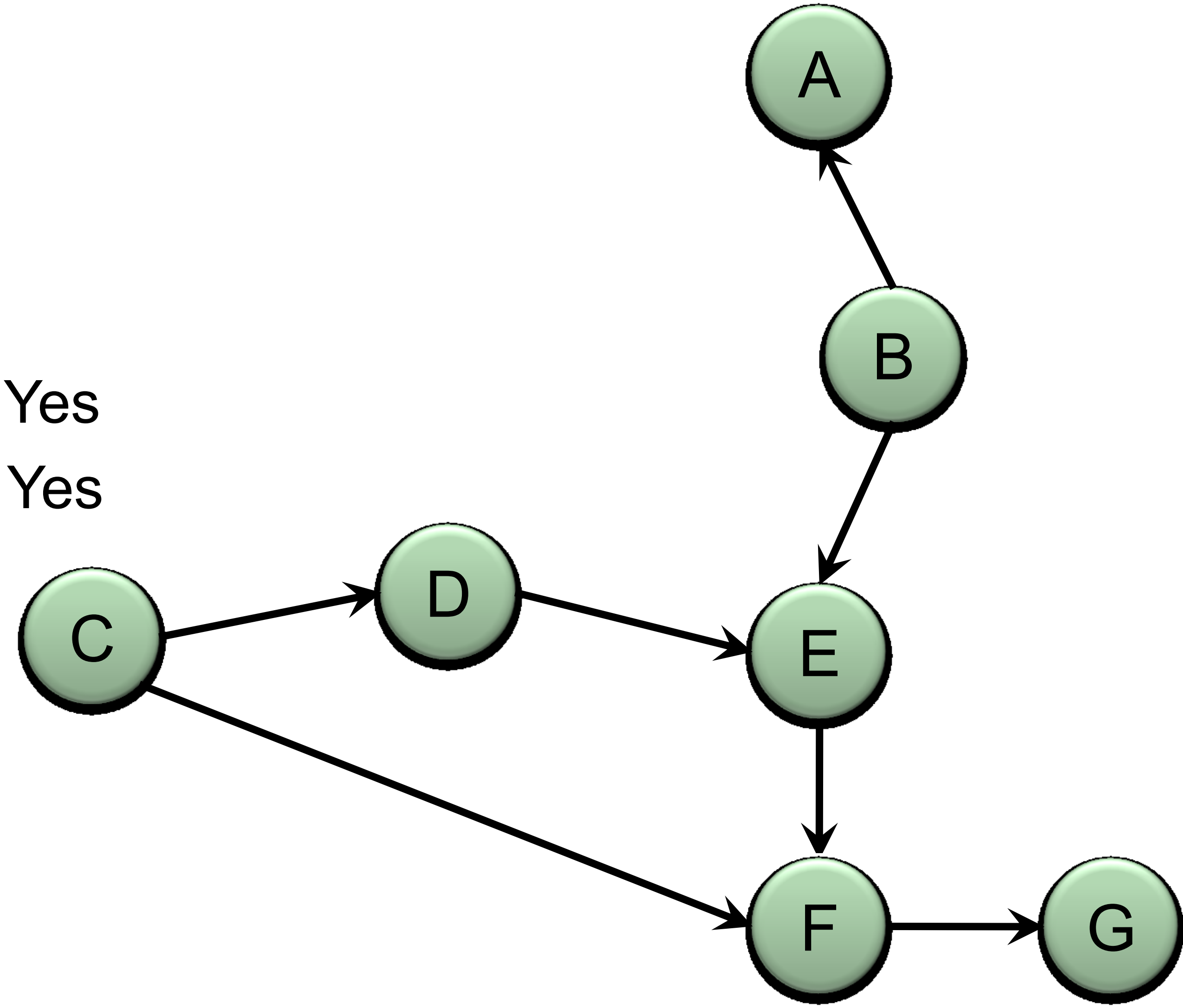
d-separation examples

- $B \perp D$?
- $B \perp D \mid E$?
- $B \perp D \mid G$?
- $D \perp F \mid C$?
- $D \perp F \mid \{C,E\}$?
- $A \perp C \mid \{B,G\}$?



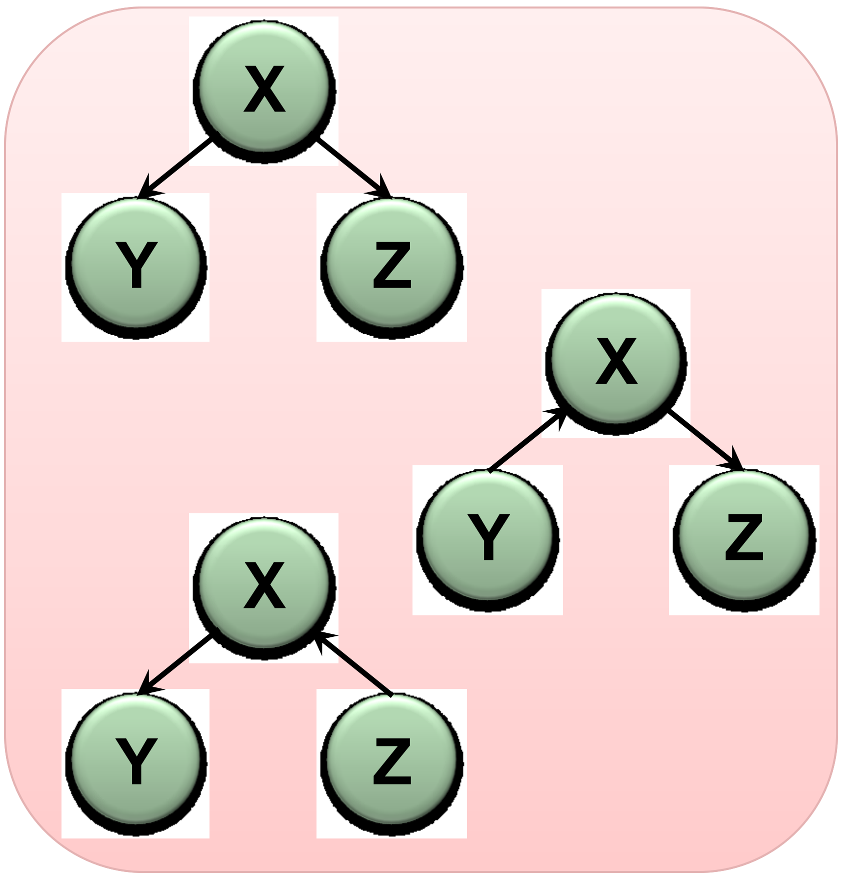
d-separation examples

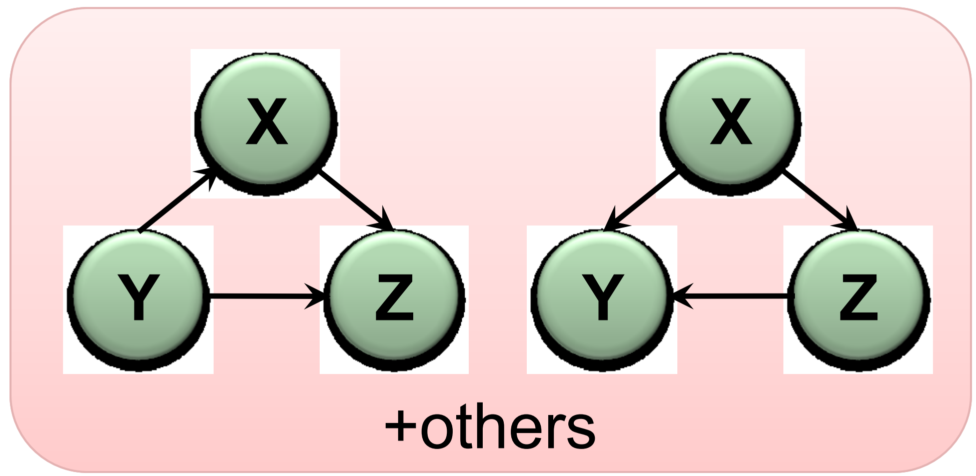
- $B \perp D$? Yes
- $B \perp D \mid E$? No
- $B \perp D \mid G$? No
- $D \perp F \mid C$? No
- $D \perp F \mid \{C,E\}$? Yes
- $A \perp C \mid \{B,G\}$? Yes

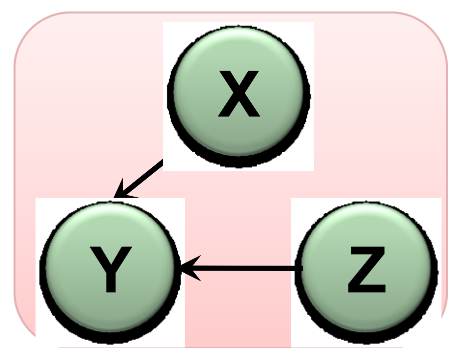


Equivalence classes

- We can group Bayes nets that have equivalent independence assumptions







Decision 2010!

Variable elimination example

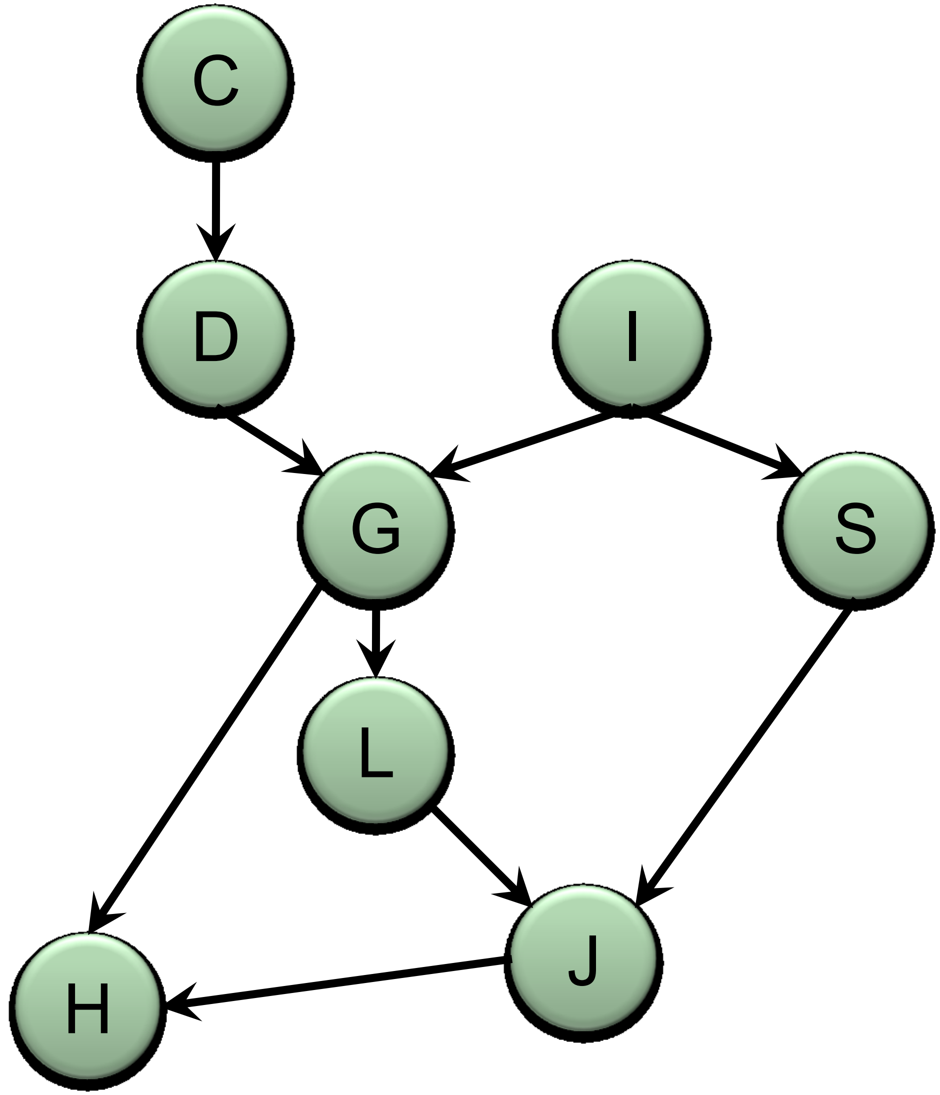
Go over midterm solutions

Contrary to what is implied above, your decision is in no way tied to the recent Pennsylvania gubernatorial election

Variable elimination example

- Student Bayes net from *Probabilistic Graphical Models* by Koller & Friedman
- Great book!

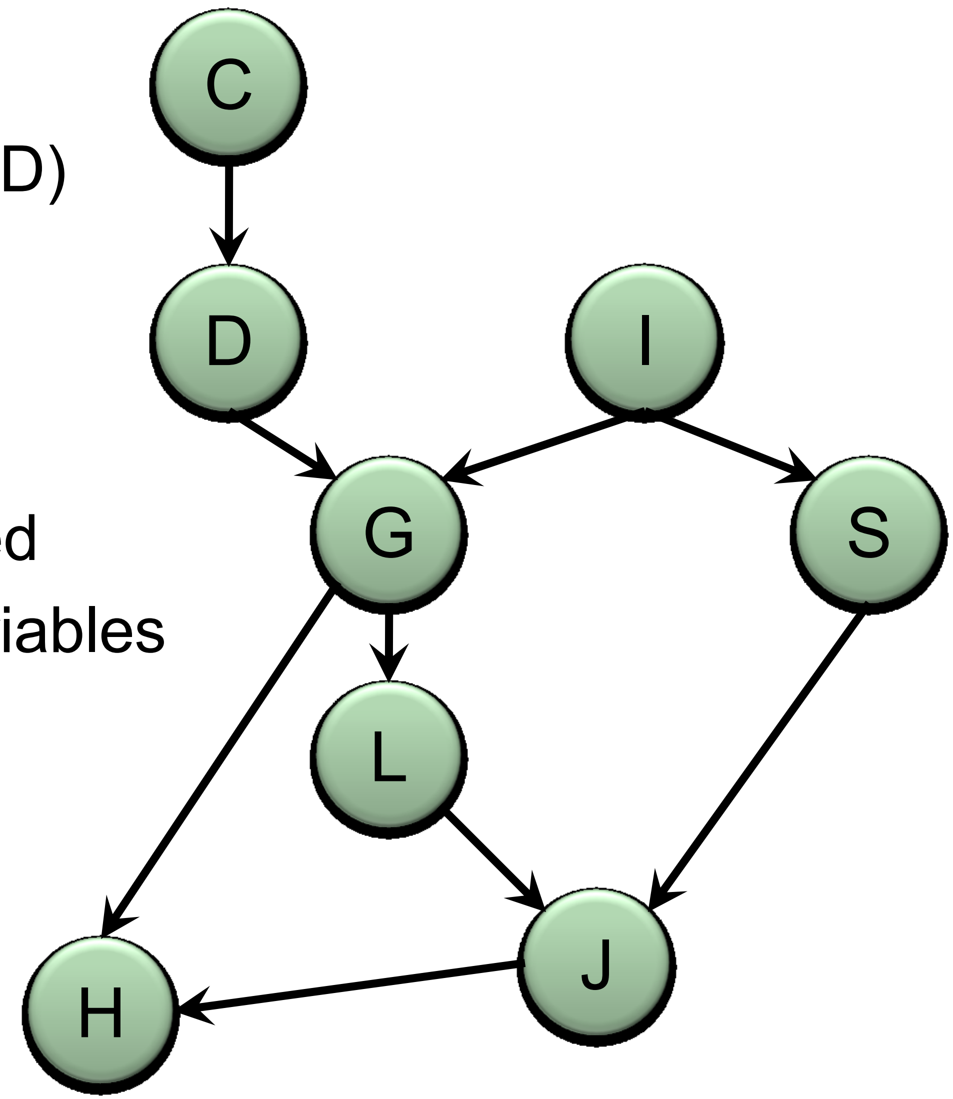
C	Coherence
D	Difficulty
I	Intelligence
G	Grade
S	SAT
L	Letter
H	Happy
J	Job



Variable elimination example

- $P(C,D,I,G,S,L,J,H) =$
  $P(C)P(D|C)P(I)P(G|I,D)$
  $P(S|I)P(L|G)P(J|L,S)$
  $P(H|G,J)$

- Want to find P(J) so need to eliminate all other variables



Variable elimination example

- Choosing an elimination order determines the order in which we sum out variables
- Use C,D,I,H,G,S,L
  - Many strategies for choosing the best order
  - It can drastically affect the efficiency
- $P(J) = \sum_L \sum_S \sum_G \sum_H \sum_I \sum_D \sum_C [P(C)P(D|C)P(I)P(G|I,D)$
  $P(S|I)P(L|G)P(J|L,S)P(H|G,J)]$

- Then pull terms out of the sums where possible
- $P(J) = \sum_L \sum_S P(J|L,S) \sum_G P(L|G) \sum_H P(H|G,J) \sum_I P(I)P(S|I)$
  $\sum_D P(G|I,D) \sum_C P(C)P(D|C)$

Variable elimination example

- $P(J) = \sum_L \sum_S P(J|L,S) \sum_G P(L|G) \sum_H P(H|G,J) \sum_I P(I)P(S|I)$
  $\sum_D P(G|I,D) \sum_C P(C)P(D|C)$

- Define functions each time a variable is summed out

  $f_1(D) = \sum_C P(C)P(D|C)$

- Values of the function taken from CPTs

  $f_1(0) = P(C=0)P(D=0|C=0) + P(C=1)P(D=0|C=1)$

  $f_1(1) = P(C=0)P(D=1|C=0) + P(C=1)P(D=1|C=1)$

- $P(J) = \sum_L \sum_S P(J|L,S) \sum_G P(L|G) \sum_H P(H|G,J) \sum_I P(I)P(S|I)$
  $\sum_D P(G|I,D) f_1(D)$

Variable elimination example

- $P(J) = \sum_L \sum_S P(J|L,S) \sum_G P(L|G) \sum_H P(H|G,J) \sum_I P(I)P(S|I)$
  $\sum_D P(G|I,D)\, f_1(D)$

- Continue defining functions – eliminate D

  $f_2(G,I) = \sum_D P(G|I,D)\, f_1(D)$
- $P(J) = \sum_L \sum_S P(J|L,S) \sum_G P(L|G) \sum_H P(H|G,J)$
  $\sum_I P(I)P(S|I)\, f_2(G,I)$

- Eliminate I

  $f_3(G,S) = \sum_I P(I)P(S|I)\, f_2(G,I)$
- $P(J) = \sum_L \sum_S P(J|L,S) \sum_G P(L|G) \sum_H P(H|G,J)\, f_3(G,S)$

Variable elimination summary

- Continue until we obtain $f_7(J)$ which is $P(J)$
- What is the advantage in doing this?
  - Functions can cache the results of each call

- What about when we are given certain variables?
  - Calculate $P(J \mid I = 0, H = 1)$
  - First calculate $f_5(J) = P(J, I = 0, H = 1)$
  - No need to sum over I and H
  - Then divide by $P(I = 0, H = 1)$
  - $P(I = 0, H = 1) = P(J=0,I=0,H=1) + P(J=1,I=0,H=1)$

    $= f_5(J=0) + f_5(J=1)$

Problem set 5

- Variable elimination question may be easiest if you type it (lots of copy/paste)
- It is okay to write a program or use a spreadsheet to do the HMM calculations
  - Do you see any similarities in the calculations for 5.3 and 5.4?
  - Still need to show the formulas you used and intermediate results
- I'm at a conference 11/16 – 11/21
  - Start early so you can ask questions
  - Watch Twitter for office hour announcements