Bayesian Networks

Lecture 7:
Bayesian Networks Continued, ICA 2

ECE/CS 498 DS
Professor Ravi K. Iyer
Department of Electrical and Computer Engineering
University of Illinois

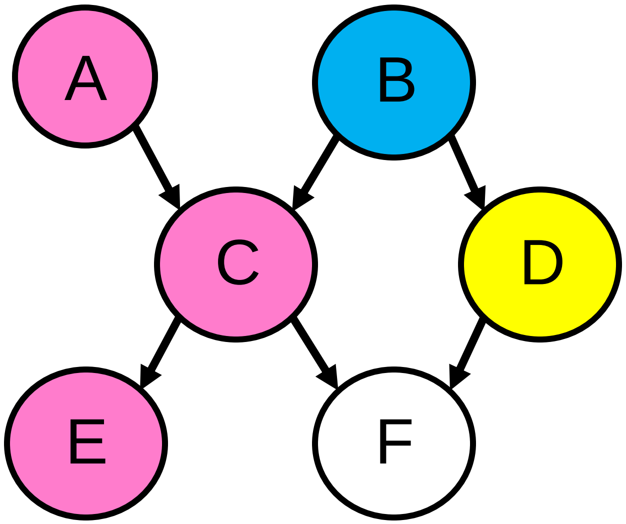
Announcements

- HW 1 due **Mon Feb 17th @ 11:59 PM** on Compass2G
  - Topic: basic Pandas review with Python
  - To be done individually
- MP 1 final checkpoint due Thu Feb 20th @ 11:59 PM on Compass2G
  - Presentations will be Fri Feb 21st in the evening
  - Presentation signups will be released by the end of the week
- Discuss section this week (2/14) will be additional practice with Bayesian Networks
- Midterm exam will take place on **Wed March 11th**

Non-Descendants: Definition

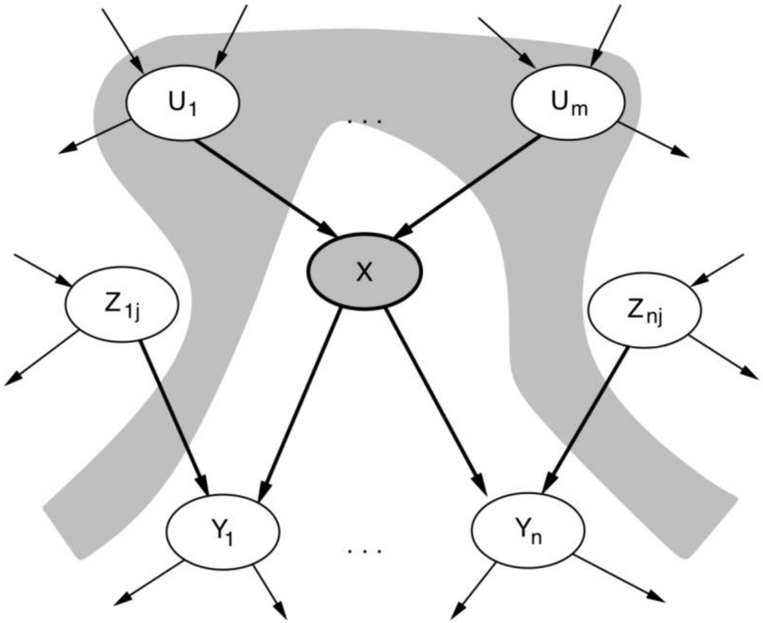
- In a directed acyclic graph (DAG) G, **non-descendants** of a node V_i are any nodes that are not descendants of V_i
- There are two types of non-descendants for a node V_i:
  - **Parents**: direct parents of V_i
  - **Other**: non-descendants that aren't parents of V_i
- For example, in the included graph,

Node	Non-Descendants		
	All	Parents	"Other"
A	B, D	None	B, D
B	A	None	A
C	A, B, D	A, B	D
$\boldsymbol{D}$	$\boldsymbol{A, B, C, E}$	$\boldsymbol{B}$	$\boldsymbol{A, C, E}$
E	A, B, C, D, F	C	A, B, D, F
F	A, B, C, D, E	C, D	A, B, E



Local Semantics

- Conditional independence was easy in NB, but for a general BN it is more involved – we will use local semantics.
- Local semantics: Each node is conditionally independent of its non-descendants given its parents
- $P(X|U_1, U_m)$ is independent of Z_{1j}, Z_{nj}
  - $P(X| U_1, U_m, Z_{1j}, Z_{nj}) = P(X| U_1, U_m)$



Local Semantics: Definition

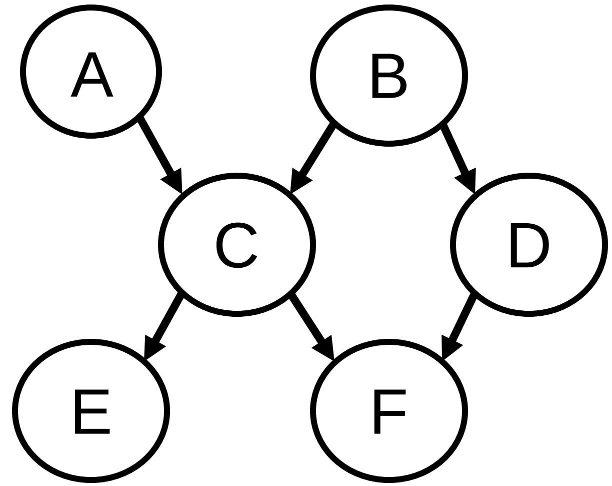
- In order to simplify calculations with joint probabilities in Bayesian Networks, one should **always condition each node on its non-descendants**

- Local Semantics for Bayesian Networks:
  - **Given its parent(s), a child is independent of its other non-descendants**
    - "Other" refers to non-descendants that aren't parents
    - <u>Alternate interpretation</u>: the probability of a node given its parents and other non-descendants is just the probability of the node given its parents
    - $\boldsymbol{P(node|parents, other\ non-descendants) = P(node|parents)}$
  - Nodes with no parents are independent of their non-descendants

- We will now go over some examples of this…

Local Semantics- Example

Consider node C in the graph to the right



- C's **non-descendants** are A, B and D, which can be further broken down into
  - **Parents**: A and B
  - **Other**: D
- Using local semantics,

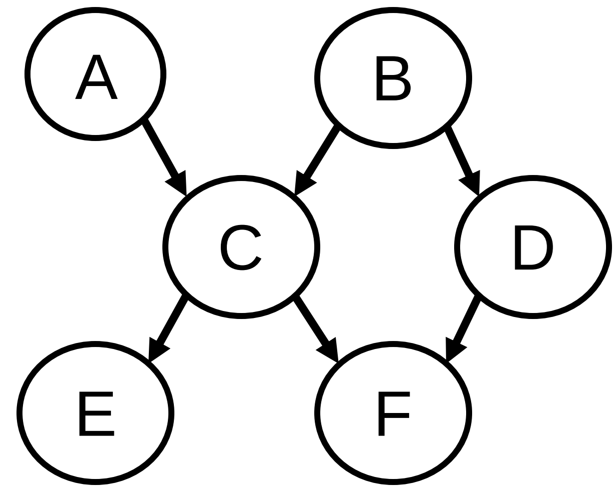
$$\boldsymbol{P(C|non-descendants) = P(C|parents)}$$

$$\Rightarrow \boldsymbol{P(C|A,B,D) = P(C|A,B)}$$

  - That is, when given its parents, C is independent of D
- Intuition:
  - Just as we learned with the Naïve Bayes classifier, ***when given the parent node, the children nodes are independent***
  - In this case, B is a common parent to both C and D. So when B is known, C and D are independent of each other
  - Thus, when B is known, D offers no additional information

Local Semantics- Example

Consider node B in the graph to the right



- B's **only non-descendant** is A
  - **Parents**: None
  - **Other**: A
- According to local semantics, since B has no parents,

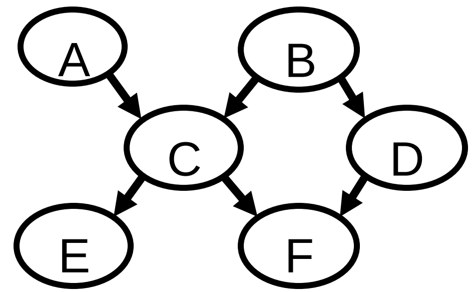
$$\boldsymbol{P(B|non-descendants) = P(B)}$$
$$\Rightarrow \boldsymbol{P(B|A) = P(B)}$$

  - That is, B is independent of its non-descendant A

- Intuition:
  - A is not a parent of B, nor vice versa
  - A and B have a common child C that, when not observed, doesn't convey information to B about A
  - Thus, knowing A offers no additional information about B

Local Semantics - Summary

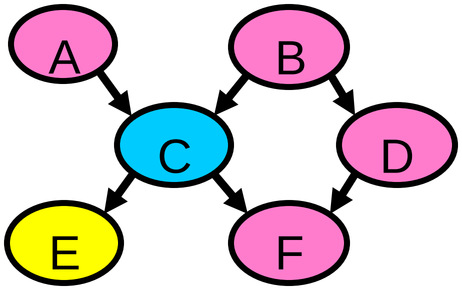
Node	Non-Descendants			Local Semantics Implication
	All	Parents	"Other"	
A	B, D	None	B, D	$P(A \mid B, D) = P(A)$
B	A	None	A	$P(B \mid A) = P(B)$
C	A, B, D	A, B	D	$P(C \mid A, B, D) = P(C \mid A, B)$
D	A, B, C, E	B	A, C, E	$P(D \mid A, B, C, E) = P(D \mid B)$
E	A, B, C, D, F	C	A, B, D, F	$P(E \mid A, B, C, D, F) = P(E \mid C)$
F	A, B, C, D, E	C, D	A, B, E	$P(F \mid A, B, C, D, E) = P(F \mid C, D)$



- **Can you explain the local semantics for nodes A, D, E, and F?**

Local Semantics - Summary

Node	Non-Descendants			Local Semantics Implication
	All	Parents	"Other"	
A	B, D	None	B, D	$P(A\|B, D) = P(A)$
B	A	None	A	$P(B\|A) = P(B)$
C	A, B, D	A, B	D	$P(C\|A, B, D) = P(C\|A, B)$
D	A, B, C, E	B	A, C, E	$P(D\|A, B, C, E) = P(D\|B)$
$\boldsymbol{E}$	$\boldsymbol{A, B, C, D, F}$	$\boldsymbol{C}$	$\boldsymbol{A, B, D, F}$	$\boldsymbol{P(E\|A, B, C, D, F) = P(E\|C)}$
F	A, B, C, D, E	C, D	A, B, E	$P(F\|A, B, C, D, E) = P(F\|C, D)$



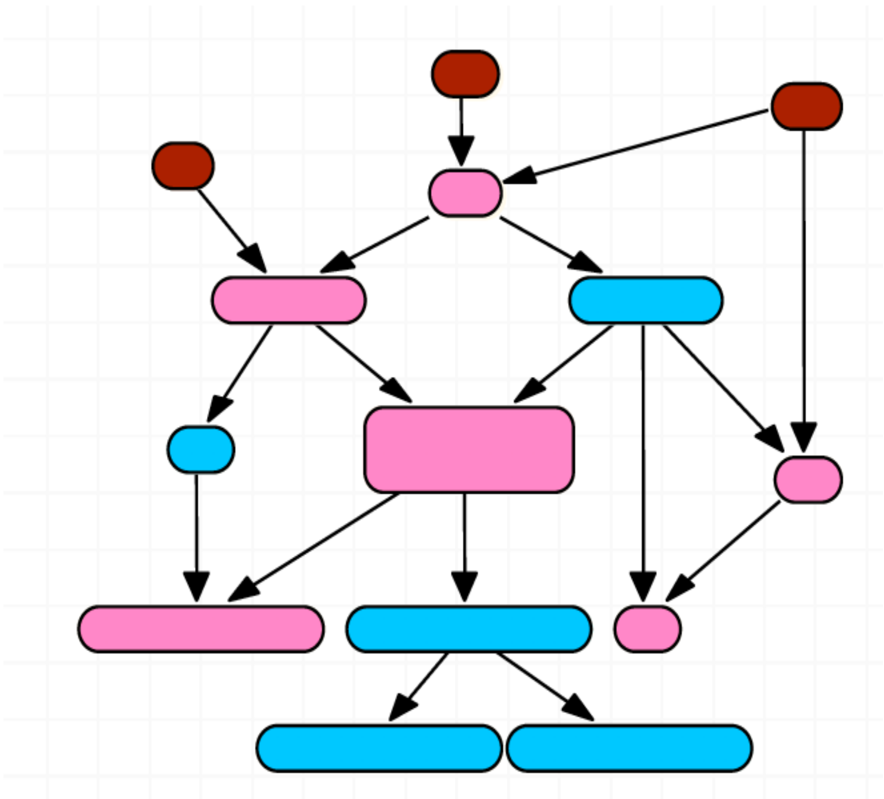
- **Can you explain the local semantics for node $\boldsymbol{E}$?**

Curse of Dimensionality

- Let Network Size = number of parameters required for joint distribution
- Network grows exponentially with number of nodes ~ 2^N
  - Each additional node doubles the size of the network!
  - A network with **100 nodes => 2^{100}-1 parameters! =>** Impractical!
- Bayesian networks can greatly reduce this complexity

Curse of Dimensionality - Example

- We are given a Bayesian network with 14 binary variables, each taking on either value 0 or 1
- The joint probability space requires 2^14 -1 ~**16 K** parameters, which is huge!
- Let us now begin to calculate the number of parameters required if we take advantage of the BN structure
  - There are 3 **red** nodes, each of which has no incoming edges and no parents
  - There are 5 **blue** nodes, each of which has 1 incoming edge from a parent
  - There are 6 **pink** nodes, each of which has 2 incoming edges from parents



Curse of Dimensionality - Example

- Consider a **red** node R. The following table defines probabilities for all values of R

$P(R = 0)$	$P(R = 1)$
k	$1 - k$

- We only need one of these two values to calculate $\mathrm{P}(R)$ since once we know one, the other is simply 1 minus the known one.
- With three red nodes, we have 3*1 = 3 total independent parameters from the red nodes

- Consider a **blue** node B with parent node P_1. The following table defines probabilities for all values of B and P_1.

$P(B\|P_1)$	$B = 0$	$B = 1$
$P_1 = 0$	k_1	$1 - k_1$
$P_1 = 1$	k_2	$1 - k_2$

- We only need one value per row to be able to calculate $\mathrm{P}(B|P_1)$, which means we need 2 independent parameters per blue node
- With 5 blue nodes, we have 5*2 = 10 total independent parameters from the blue nodes

Curse of Dimensionality - Example

- Consider a **pink** node M with parent nodes P_1 and P_2. The following table defines probabilities for all values of M, P_1 and P_2:

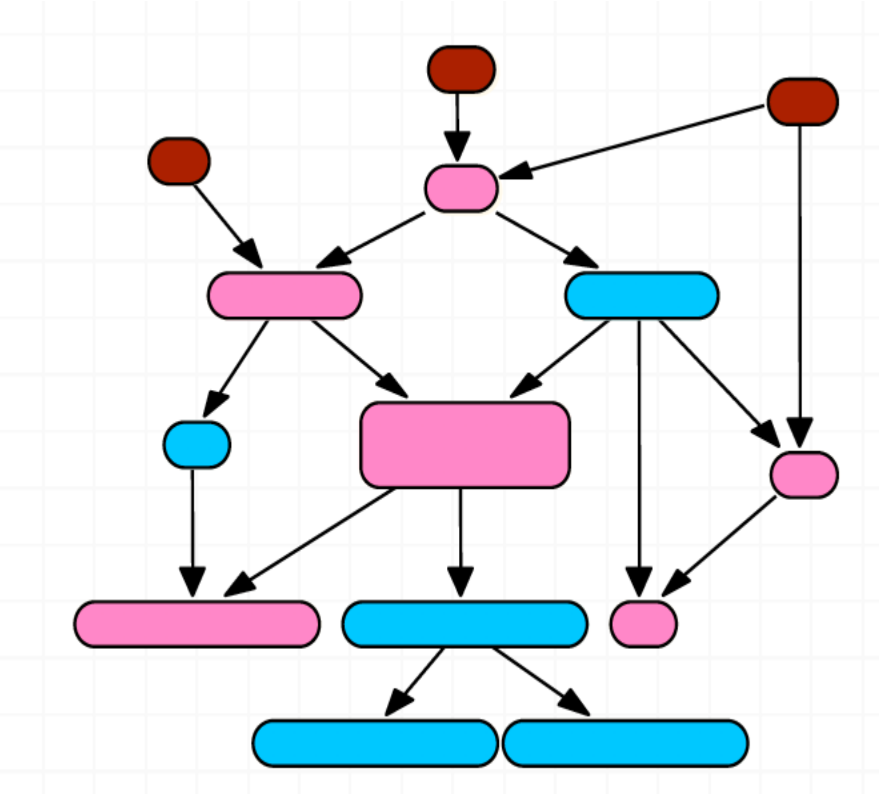
$P(M\|P_1, P_2)$	$M = 0$	$M = 1$
$P_1 = 0, P_2 = 0$	k_1	$1 - k_1$
$P_1 = 0, P_2 = 1$	k_2	$1 - k_2$
$P_1 = 1, P_2 = 0$	k_3	$1 - k_3$
$P_1 = 1, P_2 = 1$	k_4	$1 - k_4$

- We only need one value per row to be able to calculate $\mathrm{P}(M|P_1, P_2)$, which means we need 4 independent parameters per pink node
- With 6 blue nodes, we have 6*4 = 24 total independent parameters from the pink nodes

Curse of Dimensionality - Example

- We need ~**16 K** parameters to specify an arbitrary joint distribution

- Using the Bayesian Network structure as shown, we only need 3 + 10 + 24 = **37 parameters, which is much less than 16 K parameters!**



Begin ICA 2