CS 181 Spring 2021 Section 10

1 Bayesian Networks

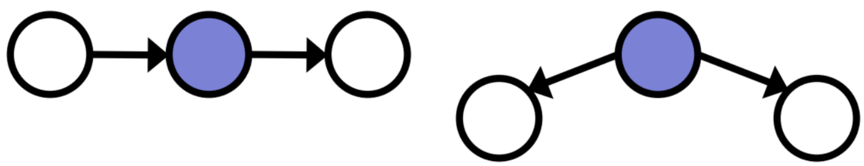
A Bayesian network is a graphical model that represents random variables and their dependencies using a directed acyclic graph. Bayesian networks are useful because they allow us to efficiently model joint distributions over many variables by taking advantage of the local dependencies. With Bayesian networks, we can easily reason about conditional independence and perform inference on large joint distributions.

1.1 D-separation rules

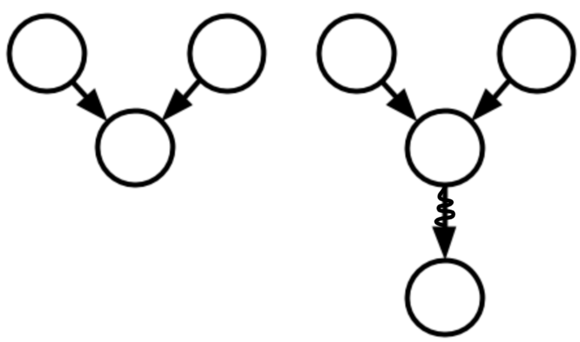
Let X_A and X_B denote sets of variables that we are interested in reasoning about. X_A and X_B are *d-separated* by a set of evidence X_E if **every** undirected path from X_A to X_B is "blocked" by X_E. A path is blocked by evidence X_E if EITHER:

1. There is a node Z with non-converging arrows on the path, and $Z \in X_E$.

   The shaded node indicates an evidence node.

2. There is a node Z with converging arrows on the path, and neither Z nor its descendants are in X_E.

Make sure to check **every** undirected path from X_A to X_B. Within each path, only one node Z needs to fall under one of the two cases described above for the whole path to be blocked.

If X_A and X_B are d-separated by X_E (i.e., all paths are blocked), then X_A and X_B are conditionally independent given X_E ($X_A \perp X_B \,|\, X_E$).

2 Network Basics

A patient goes to the doctor for a medical condition, and the doctor suspects 3 diseases as the cause of the condition. The 3 diseases are D_1, D_2, and D_3, and they are independent from each other (given no other observations). There are 4 symptoms S_1, S_2, S_3, and S_4, and the doctor wants to check for presence in order to find the most probable cause. S_1 can be caused by D_1, S_2 can be caused by D_1 and D_2, S_3 can be caused by D_1 and D_3, and S_4 can be caused by D_3. Assume all random variables are Bernoulli, i.e. the patient has the disease/symptom or not.

- **Q:** Draw a Bayesian network for this problem with the variable ordering $D_1, D_2, D_3, S_1, S_2, S_3, S_4$.

- **Q:** Write down the expression for the joint probability distribution given this network.

- **Q:** How many parameters are required to describe this joint distribution?

- **Q:** How many parameters would be required to represent the CPTs in a Bayesian network if there were no conditional independences between variables?

- **Q:** What diseases do we gain information about when observing the fourth symptom ($S_4 =$

true)?

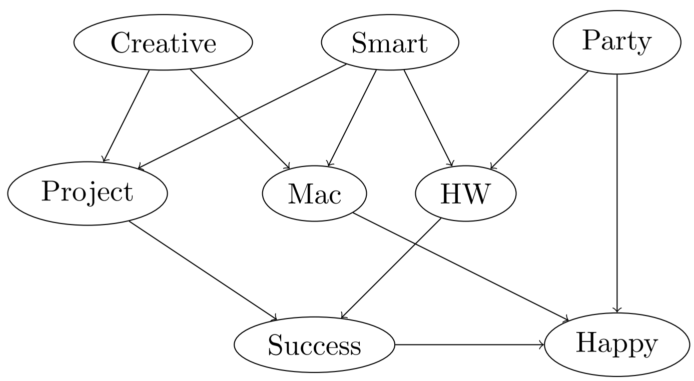
- **Q:** Suppose we know that the third symptom is present ($S_3 = true$). What does observing the fourth symptom ($S_4 = true$) tell us now?

3 D-Separation

As part of a comprehensive study of the role of CS 181 on people's happiness, we have been collecting important data from students. In an entirely optional survey that all students are required to complete, we ask the following highly objective questions:

Do you party frequently [Party: Yes/No]?
Are you smart [Smart: Yes/No]?
Are you creative [Creative: Yes/No]?
Did you do well on all your homework assignments? [HW: Yes/No]
Do you use a Mac? [Mac: Yes/No]
Did your last major project succeed? [Project: Yes/No]
Did you succeed in your most important class? [Success: Yes/No]
Are you currently Happy? [Happy: Yes/No]

After consulting behavioral psychologists we build the following model:



- **Q:** True or False: *Party* is independent of *Success* given *HW*.

- **Q:** True or False: *Creative* is independent of *Happy* given *Mac*.

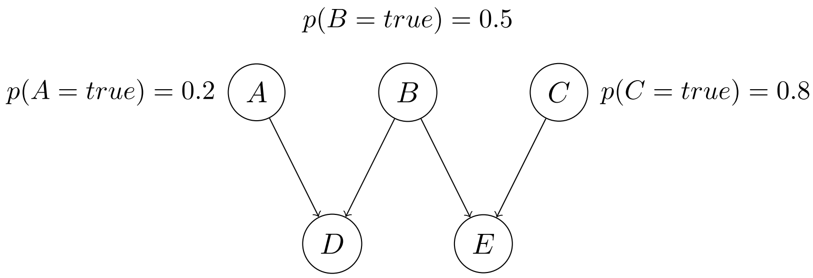
- **Q:** True or False: *Party* is independent of *Smart* given *Success*.

- **Q:** True or False: $Party$ is independent of $Creative$ given $Happy$.

- **Q:** True or False: $Party$ is independent of $Creative$ given $Success$, $Project$ and $Smart$.

4 Inference

Consider the following Bayesian network, where all variables are Bernoulli.



A	B	$p(D = true \mid A, B)$
F	F	0.9
F	T	0.6
T	F	0.5
T	T	0.1

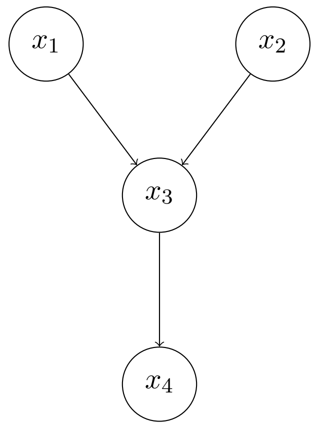
B	C	$p(E = true \mid B, C)$
F	F	0.2
F	T	0.4
T	F	0.8
T	T	0.3

- **Q:** What is the probability that all five variables are simultaneously $false$?

- **Q:** What is the probability that A is $false$ given that the remaining variables are all known to be $true$?

5 Variable Elimination in Bayesian Networks

We apply an inference algorithm called variable elimination to the following Bayesian network:



Assume that all of the random variables are Bernoulli, meaning their domain is $\{0, 1\}$ with domain size $k = 2$. In this network, we can encode the joint distribution as

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3)$$

If we wanted to calculate the marginal distribution of x_4 that is, have x_4 be our query without any evidence (conditioned on variables), we could naively marginalize out all other variables:

$$\begin{aligned} p(x_4) &= \sum_{x_1} \sum_{x_2} \sum_{x_3} p(x_1, x_2, x_3, x_4) \\ &= \sum_{x_1} \sum_{x_2} \sum_{x_3} p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3) \end{aligned}$$

To calculate these sums we would need to multiply two k-dimensional vectors for each of the $k^3 = 8$ possible combinations of x_1, x_2, x_3. In general, the number of combinations grows exponentially in the number of variables ($O(k^n)$ if you're familiar with big-O notation).

Note that Bayesian nets encode dependencies between variables, which we can use to calculate the marginal distribution more efficiently. By reordering the sums and eliminating one variable at a time, we derive the variable elimination procedure:

$$\begin{aligned} p(x_4) &= \sum_{x_1} \sum_{x_2} \sum_{x_3} p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_3) \\ &= \sum_{x_3} p(x_4|x_3) \sum_{x_2} p(x_2) \sum_{x_1} p(x_3|x_1, x_2)p(x_1) \\ &= \sum_{x_3} p(x_4|x_3) \sum_{x_2} p(x_2)p(x_3|x_2) \\ &= \sum_{x_3} p(x_4|x_3)p(x_3) \\ &= p(x_4) \end{aligned}$$

Here, we eliminate x_1 using a k by k matrix $g_1(x_3, x_2)$, because we have to sum over x_1 for each possible value of x_2 and x_3. Then we eliminate x_2 with a K-dimensional vector $g_2(x_3)$, likewise because we sum over x_2 for each possible value of x_3. Lastly, we eliminate x_3, which results in a final K-dimensional vector of probabilities for x_4. Notice that we have a poly-tree, and we're eliminating leaves first and working towards our query variable, x_4. In this way, we can perform the same computation in $O(k^2)$ time.

Alternatively, we could have eliminated variables in a different order:

$$\begin{aligned}
p(x_4) &= \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_1)p(x_2)p(x_3|x_1,x_2)p(x_4|x_3) \\
&= \sum_{x_1} p(x_1) \sum_{x_2} p(x_2) \sum_{x_3} p(x_3|x_1,x_2)p(x_4|x_3) \\
&= \sum_{x_1} p(x_1) \sum_{x_2} p(x_2)p(x_4|x_1,x_2) \\
&= \sum_{x_1} p(x_1)p(x_4|x_1) \\
&= p(x_4)
\end{aligned}$$

Here, we eliminate x_3, then x_2, then x_1. Notice that the ordering matters: eliminating x_3 first results in a $k \times k \times k$ object $g(x_1, x_2, x_4)$, so our overall algorithm will run in $O(k^3)$ time.

In general, the computational cost of variable elimination depends on the number of variables in these intermediate factors, in particular the largest object computed ('tree-width').

5.1 Exercise: Variable Elimination

Consider the Bayesian network described in above, and assume the following Conditional Probability Table (CPT). Let $x_i \in \{0, 1\}$ denote the values that variable X_i can take. Our goal is to find $p(x_4)$.

x_1	$p(x_1)$
0	0.3
1	0.7

x_2	$p(x_2)$
0	0.6
1	0.4

x_3	x_1	x_2	$p(x_3\|x_1, x_2)$
0	0	0	0.5
0	0	1	0.2
0	1	0	0.9
0	1	1	0.5
1	0	0	0.5
1	0	1	0.8
1	1	0	0.1
1	1	1	0.5

x_4	x_3	$p(x_4\|x_3)$
0	0	0.7
0	1	0.1
1	0	0.3
1	1	0.9

1. Eliminate X_1 first. Draw the resulting Bayesian network and compute the CPT.
2. Eliminate X_3 first. Draw the resulting Bayesian network and compute the CPT.
3. How many sum-product calculations do each of these variable elimination orders require? Which one is preferable?