CS 181 Spring 2020 Section 8
Solution

1 Bayesian Networks

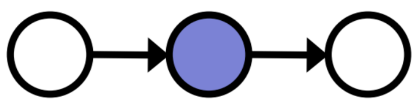
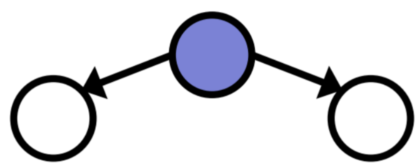
A Bayesian network is a graphical model that represents random variables and their dependencies using a directed acyclic graph. Bayesian networks are useful because they allow us to efficiently model joint distributions over many variables by taking advantage of the local dependencies. With Bayesian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

1.1 D-separation rules

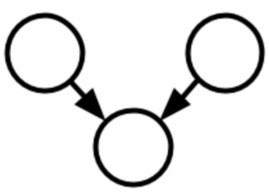
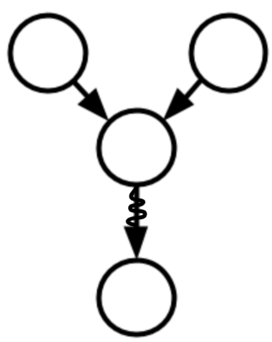
Let X_A and X_B denote sets of variables that we are interested in reasoning about. X_A and X_B are *d-separated* by a set of evidence X_E if **every** undirected path from X_A to X_B is "blocked" by X_E. A path is blocked by evidence X_E if EITHER:

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.

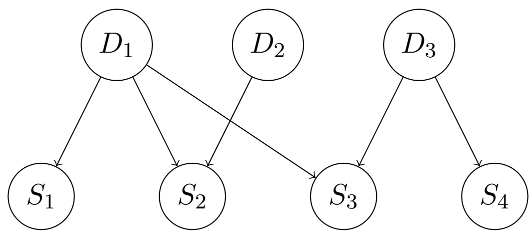
If X_A and X_B are d-separated by X_E (i.e., blocked), then X_A and X_B are conditionally independent given X_E ($X_A \perp X_B \mid X_E$).

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 $D_1, D_2, D_3, S_1, S_2, S_3, S_4$.

  A: Note that there are many valid networks (depending on the chosen variable ordering), some more efficient (i.e. requiring fewer parameters) than others. Here is a compact representation that comes from variable ordering $D_1, D_2, D_3, S_1, S_2, S_3, S_4$. (Recall that all dependencies to earlier variables need to be indicated with edges).



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

  A: $p(D_1, D_2, D_3, S_1, S_2, S_3, S_4)$

  $= p(D_1)p(D_2)p(D_3)p(S_1|D_1)p(S_2|D_1, D_2)p(S_3|D_1, D_3)p(S_4|D_3)$

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

  A:

Conditional Probability Table	Number of Parameters
$p(D_1)$	1
$p(D_2)$	1
$p(D_3)$	1
$p(S_1\|D_1)$	2
$p(S_2\|D_1, D_2)$	4
$p(S_3\|D_1, D_3)$	4
$p(S_4\|D_3)$	2
Total Number of Parameters	15

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

  A: The network would be structured as a clique, and considering order $D_1, D_2, D_3, S_1, S_2, S_3, S_4$, the number of parameters for the CPTs would be $1 + 2 + 4 + 8 + 16 + 32 + 64 = 127$.

Conditional Probability Table	Number of Parameters
$p(D_1)$	1
$p(D_2\|D_1)$	2
$p(D_3\|D_1, D_2)$	4
$p(S_1\|D_1, D_2, D_3)$	8
$p(S_2\|D_1, D_2, D_3, S_1)$	16
$p(S_3\|D_1, D_2, D_3, S_1, S_2)$	32
$p(S_4\|D_1, D_2, D_3, S_1, S_2, S_3)$	64
Total Number of Parameters	127

(We can see there is no saving relative to specifying the joint probability distribution directly, which would require $2^7 - 1 = 127$ numbers.)

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

  A: We have independence relations $I(D_1, S_4)$ (since the path is blocked without observing S_3 and $I(D_2, S_4)$ (since the path is blocked at both S_2 and S_3). What is left is dependence between D_3 and S_4. Thus, we only learn information about D_3.

- **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?

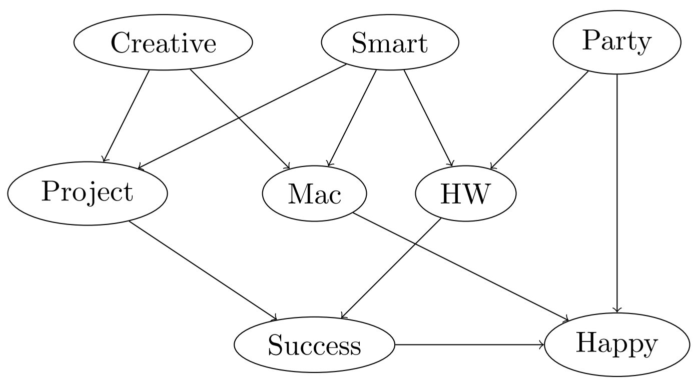
  A: With $S_3 = true$, observing $S_4 = true$ now also gives us information about D_1 (via 'explaining away', or using d-separation, because the D_1 to S_4 path is no longer blocked at S_3). We still don't learn any information about D_2 because the D_2 to S_4 path remains blocked at S_2.

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*.

  A: False; there is a path that is not blocked: *Party* – *HW* – *Smart* – *Project* – *Success* has neither a converging arrows not in the set of evidence or a non-converging arrows in the set.

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

  A: False; there is a path that is not blocked: *Creative* – *Project* – *Success* – *Happy*

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

  A: False; there is a path that is not blocked between *Party* and *Smart*: the path *Party* – *HW* – *Success* is not blocked because the converging arrows node at *HW* has a descendant (*Success*) in the evidence.

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

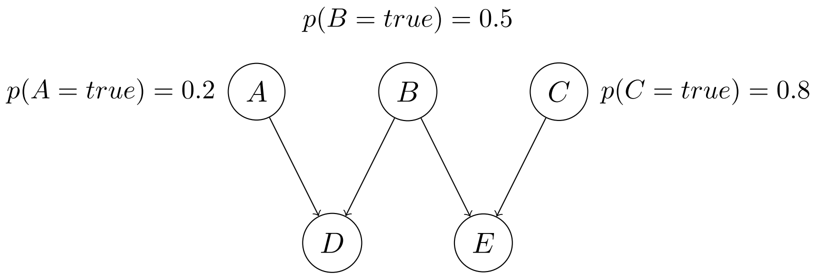
A: False; there is a path that is not blocked between *Party* and *Creative* through the converging arrows at *Happy*. There are actually multiple not-blocked paths – can you find them?

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

  A: True! All paths between *Party* and *Creative* are blocked. Working from *Party*, the paths that come through *Happy* are blocked there (converging arrows, no evidence). Those that come through *HW* and *Smart* are blocked at *Smart*. Those that come through *HW*, *Success*, *Project* are blocked at *Project*.

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 $false$?

  A:

$$
\begin{aligned}
p(\neg A, \neg B, \neg C, \neg D, \neg E) &= p(\neg A)p(\neg B)p(\neg C)p(\neg D|\neg A, \neg B)p(\neg E|\neg B, \neg C) \\
&= (0.8)(0.5)(0.2)(0.1)(0.8) \\
&= 0.0064
\end{aligned}
$$

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

  A: For this part, we need to calculate $p(\neg A|B, C, D, E)$.

  By the definition of conditional probability,

$$
p(\neg A|B, C, D, E) = \frac{p(\neg A, B, C, D, E)}{P(B, C, D, E)} = \frac{p(\neg A, B, C, D, E)}{P(\neg A, B, C, D, E) + P(A, B, C, D, E)}
$$

  The joint probabilities $p(\neg A, B, C, D, E)$ and $p(A, B, C, D, E)$ can be computed as:

$$
\begin{aligned}
p(\neg A, B, C, D, E) &= p(\neg A)p(B)p(C)p(D|\neg A, B)p(E|B, C) \\
&= (0.8)(0.5)(0.8)(0.6)(0.3) \\
&= (0.05760) \\
p(A, B, C, D, E) &= p(A)p(B)p(C)p(D|A, B)p(E|B, C) \\
&= (0.2)(0.5)(0.8)(0.1)(0.3) \\
&= (0.00240)
\end{aligned}
$$

Finally, we can plug this in to get:

$$p(\neg A|B, C, D, E) = \frac{.05760}{.05760 + .00240} = .96$$