6.825 Techniques in Artificial Intelligence

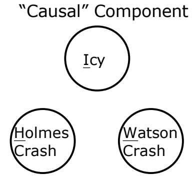
Bayesian Networks

- To do probabilistic reasoning, you need to know the joint probability distribution
- But, in a domain with N propositional variables, one needs 2^N numbers to specify the joint probability distribution
- We want to exploit independences in the domain
- Two components: structure and numerical parameters

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

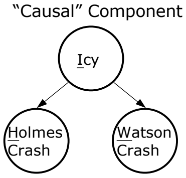
"Causal" Component



Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

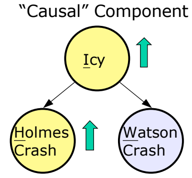
"Causal" Component



Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

"Causal" Component

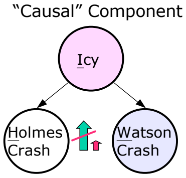


H and W are dependent,

Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

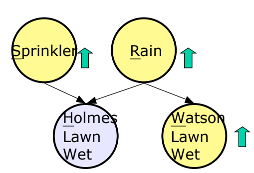
"Causal" Component



H and W are dependent, but conditionally independent given I

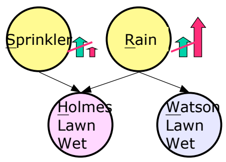
Holmes and Watson in LA

Holmes and Watson have moved to LA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



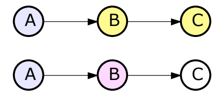
Holmes and Watson in LA

Holmes and Watson have moved to LA. He wakes up to find his lawn wet. He wonders if it has rained or if he left his sprinkler on. He looks at his neighbor Watson's lawn and he sees it is wet too. So, he concludes it must have rained.



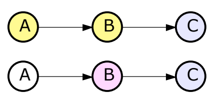
Given W, P(R) goes up and P(S) goes down – "explaining away"

Forward Serial Connection



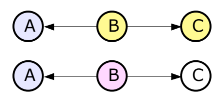
- Transmit evidence from A to C through unless B is instantiated (its truth value is known)
  - A = battery dead
  - B = car won't start
  - C = car won't move
- Knowing about A will tell us something about C
- But, if we know B, then knowing about A will not tell us anything new about C.

Backward Serial Connection



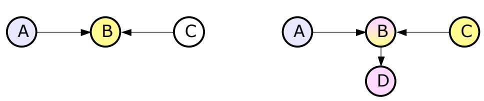
- Transmit evidence from C to A through unless B is instantiated (its truth value is known)
  - A = battery dead
  - B = car won't start
  - C = car won't move
- Knowing about C will tell us something about A
- But, if we know B, then knowing about C will not tell us anything new about A

Diverging Connection



- Transmit evidence through B unless it is instantiated
  - A = Watson crash
  - B = Icy
  - C = Holmes crash
- Knowing about A will tell us something about C
- Knowing about C will tell us something about A
- But, if we know B, then knowing about A will not tell us anything new about C, or vice versa

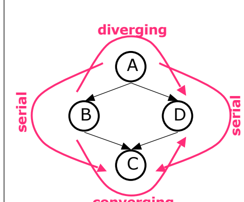
Converging Connection



- Transmit evidence from A to C only if B or a descendant of B is instantiated
  - A = Bacterial infection
  - B = Sore throat
  - C = Viral Infection
- Without knowing B, finding A does not tell us anything about B
- If we see evidence for B, then A and C become dependent (potential for "explaining away"). If we find bacteria in patient with a sore throat, then viral infection is less likely.

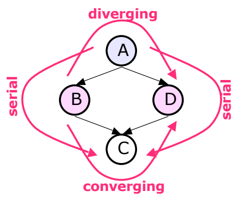
D-separation

- Two variables A and B are d-separated iff for every path between them, there is an intermediate variable V such that either
  - The connection is serial or diverging and V is known
  - The connection is converging and neither V nor any descendant is instantiated
  - Two variables are d-connected iff they are not d-separated



- A-B-C: serial, blocked when B is known, connected otherwise
- A-D-C: serial, blocked when D is known, connected otherwise
- B-A-D: diverging, blocked when A is known, connected otherwise
- B-C-D: converging, blocked when C has no evidence, connected otherwise

D-Separation Detail

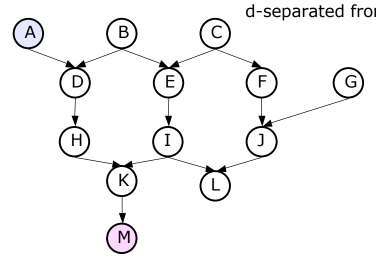


- A-B-C: serial, blocked when B is known, connected otherwise
- A-D-C: serial, blocked when D is known, connected otherwise
- B-A-D: diverging, blocked when A is known, connected otherwise
- B-C-D: converging, blocked when C has no evidence, connected otherwise

- No instantiation
  - A, C are d-connected (A-B-C connected, A-D-C connected)
  - B, D are d-connected (B-A-D connected, B-C-D blocked)
- A instantiated
  - B, D are d-separated (B-A-D blocked, B-C-D blocked)
- A and C instantiated
  - B, D are d-connected (B-A-D blocked, B-C-D connected)
- B instantiated
  - A, C are d-connected (A-B-C blocked, A-D-C connected)
- B and D instantiated
  - A, C are d-separated (A-B-C blocked, A-D-C blocked)

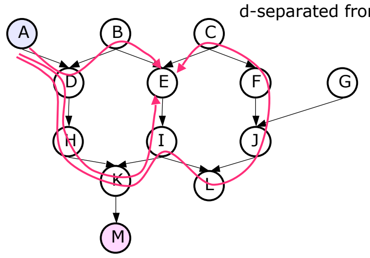
D-Separation Example

Given M is known, is A d-separated from E?



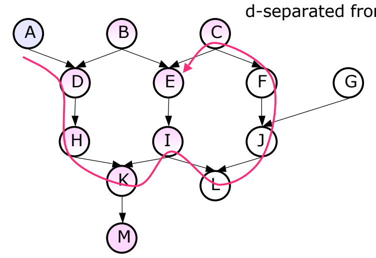
D-Separation Example

Given M is known, is A d-separated from E?



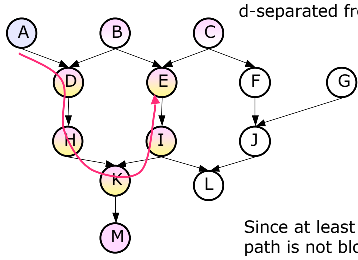
D-Separation Example

Given M is known, is A d-separated from E?



D-Separation Example

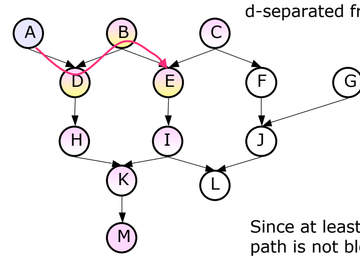
Given M is known, is A d-separated from E?



Since at least one path is not blocked, A is not d-separated from E

D-Separation Example

Given M is known, is A d-separated from E?



Since at least one path is not blocked, A is not d-separated from E

Recitation Problems

Use the Bayesian network from the previous slides to answer the following questions:

- Are A and F d-separated if M is instantiated?
- Are A and F d-separated if nothing is instantiated?
- Are A and E d-separated if I is instantiated?
- Are A and E d-separated if B and H are instantiated?
- Describe a situation in which A and G are d-separated.
- Describe a situation in which A and G are d-connected.

Bayesian (Belief) Networks

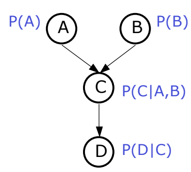
- Set of variables, each has a finite set of values
- Set of directed arcs between them forming acyclic graph
- Every node A, with parents $B_1, \ldots, B_n$, has $P(A \mid B_1,\ldots,B_n)$ specified

Theorem: If A and B are d-separated given evidence e, then P(A | e) = P(A | B, e)

Chain Rule

- Variables: $V_1, \ldots, V_n$
- Values: $v_1, \ldots, v_n$
- $P(V_1=v_1, V_2=v_2, \ldots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

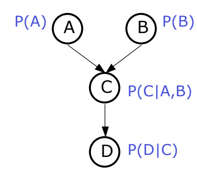
P(ABCD) = P(A=true, B=true, C=true, D=true)



Chain Rule

- Variables: $V_1, \ldots, V_n$
- Values: $v_1, \ldots, v_n$
- $P(V_1=v_1, V_2=v_2, \ldots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

P(ABCD) = P(A=true, B=true, C=true, D=true)



P(ABCD) =

P(D|ABC)P(ABC) =

P(D|C) P(ABC) =

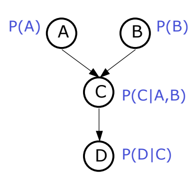
A d-separated from D given C

B d-separated from D given C

Chain Rule

- Variables: $V_1, \ldots, V_n$
- Values: $v_1, \ldots, v_n$
- $P(V_1=v_1, V_2=v_2, \ldots, V_n=v_n) = \prod_i P(V_i=v_i \mid \text{parents}(V_i))$

P(ABCD) = P(A=true, B=true, C=true, D=true)



P(ABCD) =

P(D|ABC)P(ABC) =

P(D|C) P(ABC) =

P(D|C) P(C|AB) P(AB) =

P(D|C) P(C|AB) P(A)P(B)

A d-separated from D given C

B d-separated from D given C

A d-separated from B

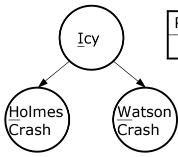
Key Advantage

- The conditional independencies (missing arrows) mean that we can store and compute the joint probability distribution more efficiently

Icy Roads with Numbers

t = true

f = false



P(I=t)	P(I=f)
0.7	0.3

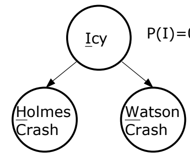
	P(H=t | I)	P(H=f | I)
I=t	0.8	0.2
I=f	0.1	0.9

	P(W=t | I)	P(W=f | I)
I=t	0.8	0.2
I=f	0.1	0.9

The right-hand column in these tables is redundant, since we know the entries in each row must add to 1.

NB: the columns need NOT add to 1.

Numerical Example: Shorthand

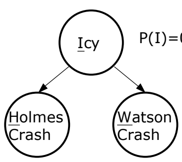


P(I)=0.7

	P(H| I)
I	0.8
¬I	0.1

	P(W| I)
I	0.8
¬I	0.1

Probability that Watson Crashes



P(I)=0.7

	P(H| I)
I	0.8
¬I	0.1

	P(W| I)
I	0.8
¬I	0.1

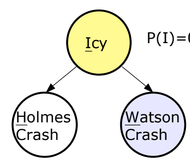
P(W) = P(W| I) P(I) + P(W| ¬I) P(¬I)

= 0.8 0.7 + 0.1 0.3

= 0.56 + 0.03

= 0.59

Probability of Icy given Watson



P(I)=0.7

	P(H| I)
I	0.8
¬I	0.1

	P(W| I)
I	0.8
¬I	0.1

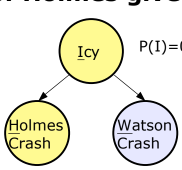
P(I | W) = P(W | I) P(I) / P(W)

= 0.8 0.7 / 0.59

= 0.95

We started with P(I) = 0.7; knowing that Watson crashed raised the probability to 0.95

Probability of Holmes given Watson



P(I)=0.7

	P(H| I)
I	0.8
¬I	0.1

	P(W| I)
I	0.8
¬I	0.1

P(H|W) = P(H|W,I)P(I|W) + P(H|W,¬I) P(¬I| W)

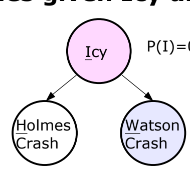
= P(H|I)P(I|W) + P(H|¬I) P(¬I| W)

= 0.8 0.95 + 0.1 0.05

= 0.765

We started with P(H) = 0.59; knowing that Watson crashed raised the probability to 0.765

Prob of Holmes given Icy and Watson



P(I)=0.7

	P(H| I)
I	0.8
¬I	0.1

	P(W| I)
I	0.8
¬I	0.1

P(H|W, ¬I) = P(H|¬I) = 0.1

H and W are d-separated given I, so H and W are conditionally independent given I

Recitation Problems II

In the Watson and Holmes visit LA network, use the following conditional probability tables.

P(R) = 0.2

P(S) = 0.1

	P(W| R)
R	1.0
¬R	0.2

	P(H| R,S)
R,S	1.0
R,←¬S	1.0
¬ R,←S	0.9
¬ R,←¬S	0.1

Calculate:

P(H), P(R|H), P(S|H), P(W|H), P(R|W,H), P(S|W,H)