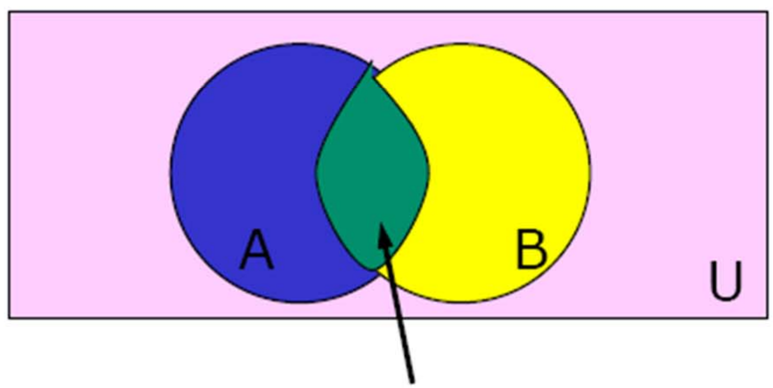
Introduction to Bayesian Networks

Slides based on course of Leslie Pack Kaelbling, MIT

Axioms of Probability

- Universe of atomic events (like interpretations in logic).
- Events are sets of atomic events
- P: events $\rightarrow$ [0,1]
  - P(true) = 1 = P(U)
  - P(false) = 0 = P($\emptyset$)
  - P(A v B) = P(A) + P(B) – P(A Æ B)



A Æ B

A Question

Jane is from Berkeley. She was active in anti-war protests in the 60's. She lives in a commune.

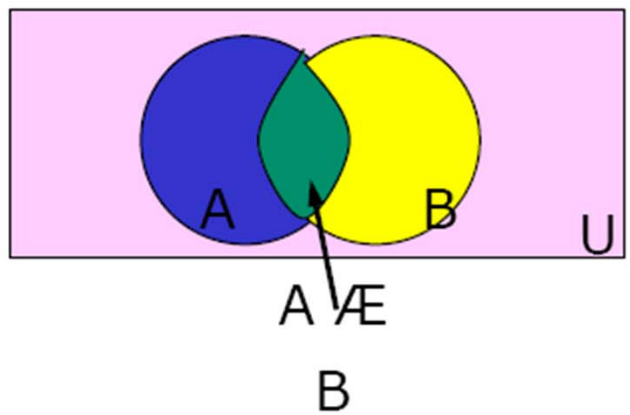
- Which is more probable?
  1. Jane is a bank teller
  2. Jane is a feminist bank teller

A Question

Jane is from Berkeley. She was active in anti-war protests in the 60's. She lives in a commune.

- Which is more probable?
  1. Jane is a bank teller
  2. Jane is a feminist bank teller

1. A
2. A Æ B



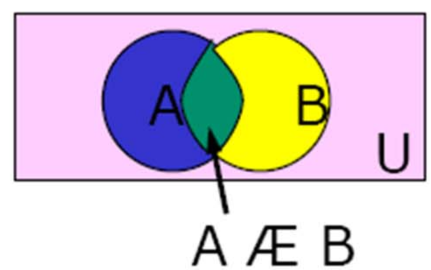
Random Variables

- Random variables
  - Function: discrete domain $\rightarrow$ [0, 1]
  - Sums to 1 over the domain
    - Raining is a propositional random variable
    - Raining(true) = 0.2
      - P(Raining = true) = 0.2
    - Raining(false) = 0.8
      - P(Raining = false) = 0.8
- Joint distribution
  - Probability assignment to all combinations of values of random variables

Joint Distribution Example

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

- The sum of the entries in this table has to be 1
- Given this table, one can answer all the probability questions about this domain
- P(cavity) = 0.1 [add elements of cavity row]
- P(toothache) = 0.05 [add elements of toothache column]
- P(A | B) = P(A Æ B)/P(B) [prob of A when U is limited to B]
- P(cavity | toothache) = 0.04/0.05 = 0.8



Independence

- A and B are independent iff
  - $P(A \text{ Æ } B) = P(A) \cdot P(B)$
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$
- Independence is essential for efficient probabilistic reasoning
- A and B are conditionally independent given C iff
  - $P(A \mid B, C) = P(A \mid C)$
  - $P(B \mid A, C) = P(B \mid C)$
  - $P(A \text{ Æ } B \mid C) = P(A \mid C) \cdot P(B \mid C)$

Examples of Conditional Independence

- Toothache (T)
- Spot in Xray (X)
- Cavity (C)
- None of these propositions are independent of one other
- T and X are conditionally independent given C

- Battery is dead (B)
- Radio plays (R)
- Starter turns over (S)
- None of these propositions are independent of one another
- R and S are conditionally independent given B

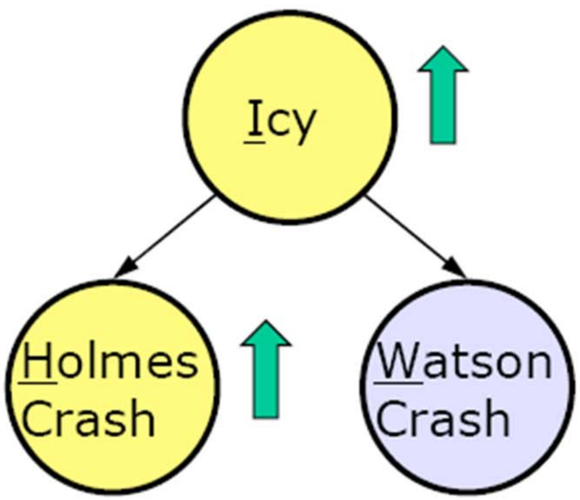
Bayesian Networks

- 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

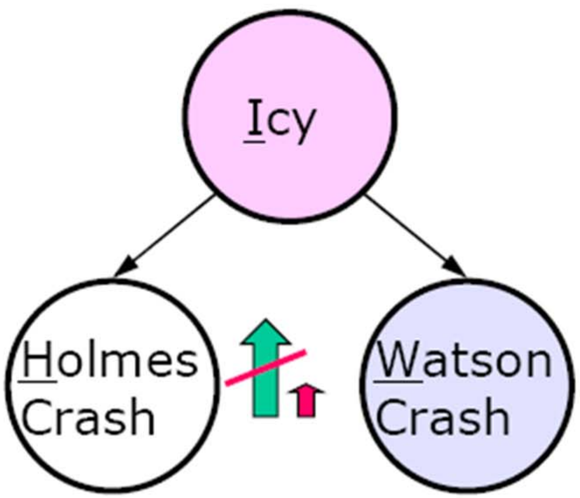


H and W are dependent,

Icy Roads

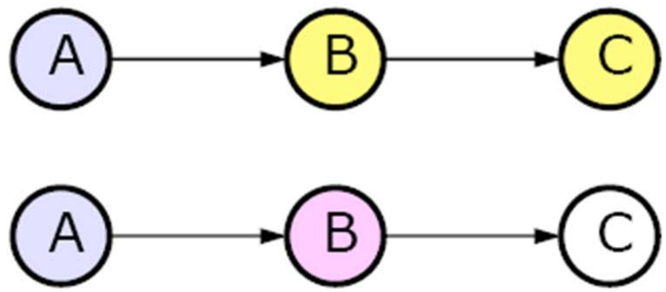
Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, “It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch.” But, his secretary says,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

“Causal” Component



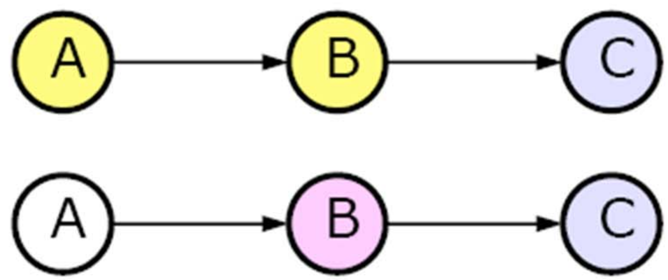
H and W are dependent, but conditionally independent given I

Forward Serial Connection



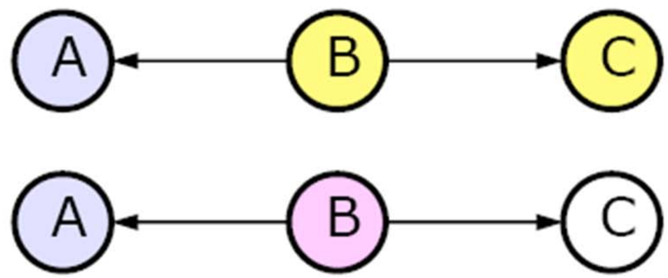
- Transmit evidence from A to C through unless B is instantiated (its truth value is known)
    - A = battery dead
    - B = car won’t start
    - C = car won’t move
- Knowing about A will tell us something about C
- But, if we know B, then knowing about A will not tell us anything new about C.

Backward Serial Connection



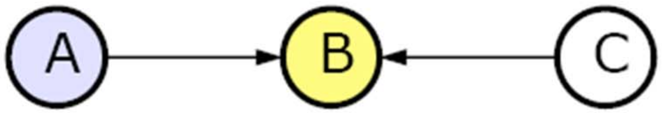
- Transmit evidence from C to A through unless B is instantiated (its truth value is known)
  - A = battery dead
  - B = car won’t start
  - C = car won’t move
- Knowing about C will tell us something about A
- But, if we know B, then knowing about C will not tell us anything new about A

Diverging Connection



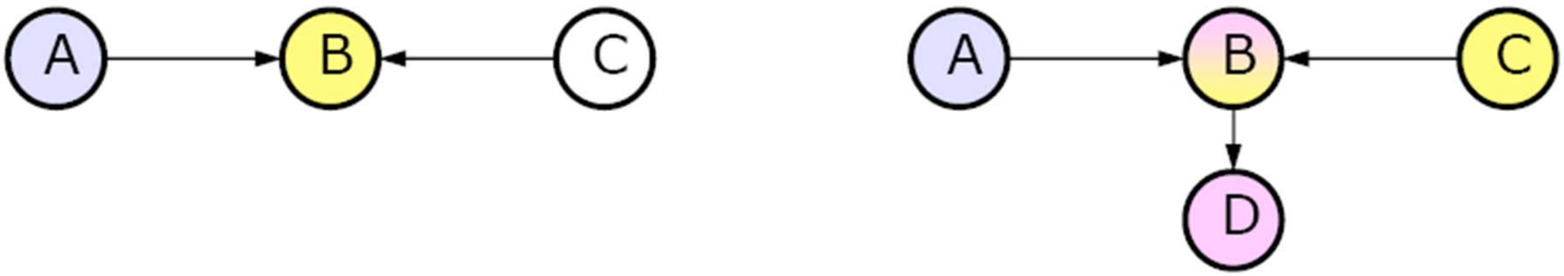
- Transmit evidence through B unless it is instantiated
  - A = Watson crash
  - B = Icy
  - C = Holmes crash
- Knowing about A will tell us something about C
- Knowing about C will tell us something about A
- But, if we know B, then knowing about A will not tell us anything new about C, or vice versa

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

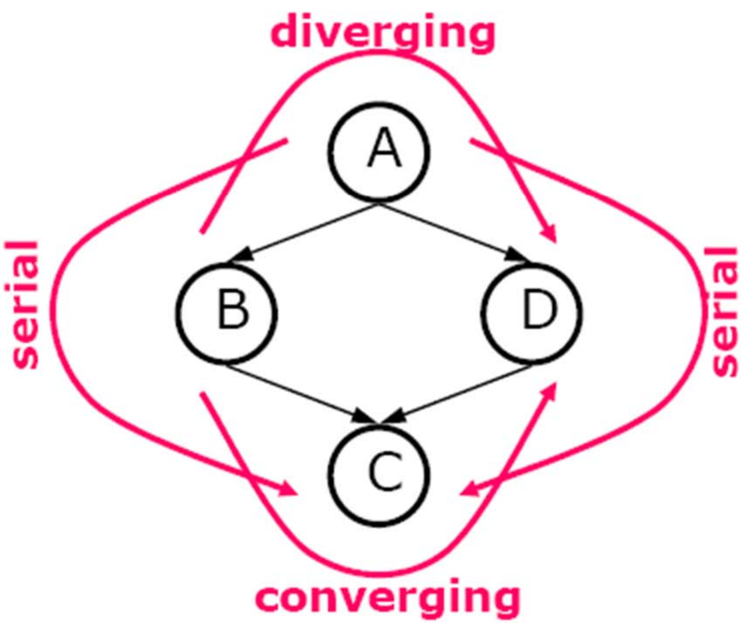
Converging Connection



- 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.

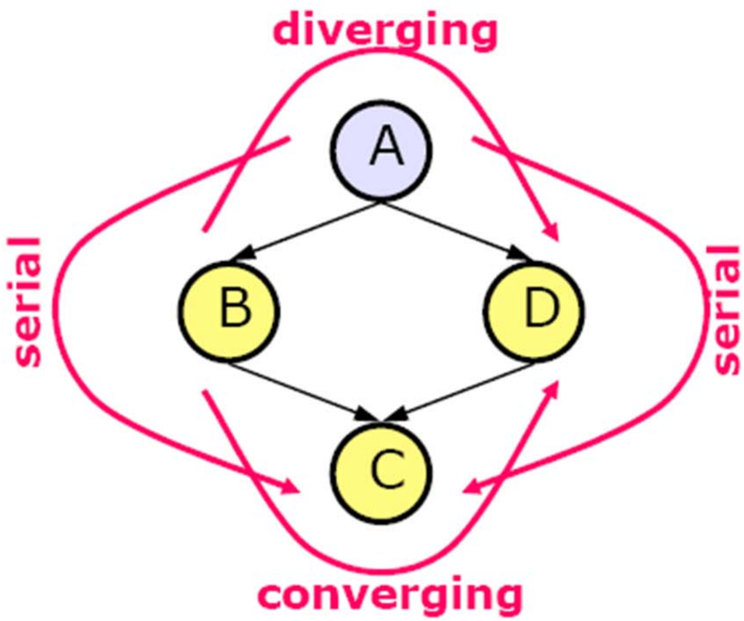
D-separation

- 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

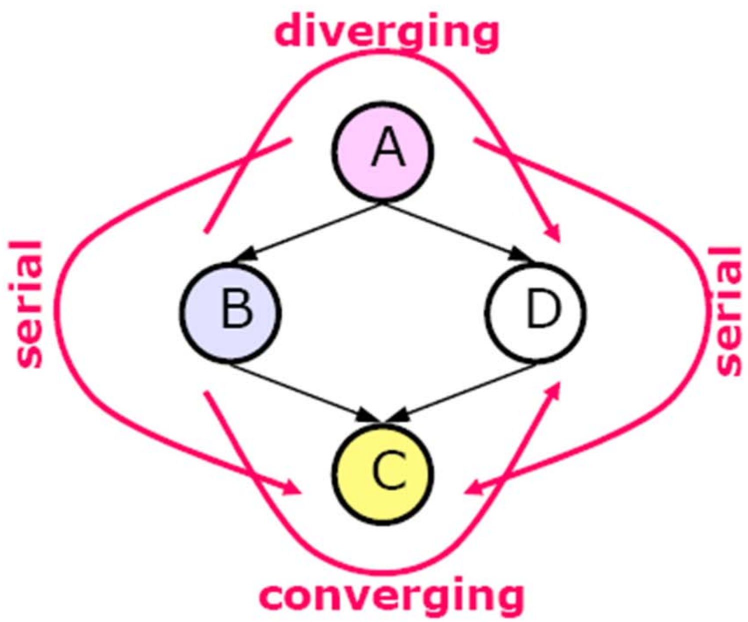
D-Separation Detail



- No instantiation
  - A, C are d-connected (A-B-C connected, A-D-C connected)

- 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

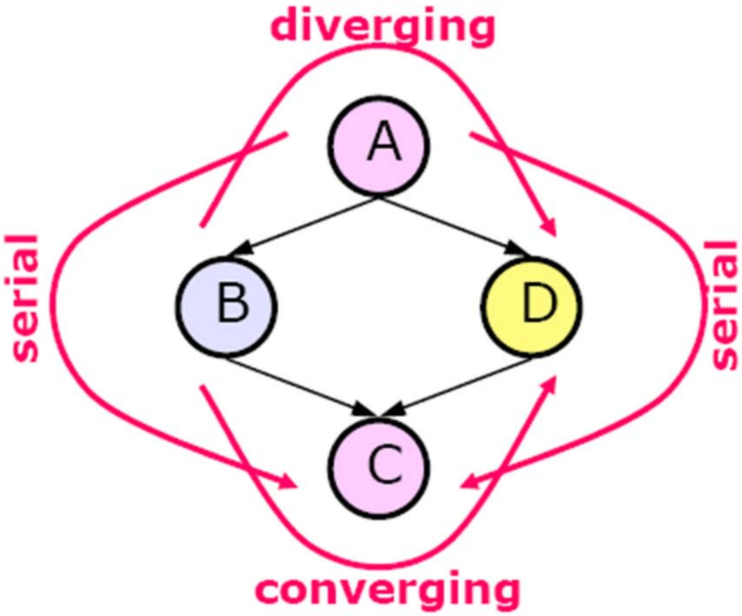
D-Separation Detail



- 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-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

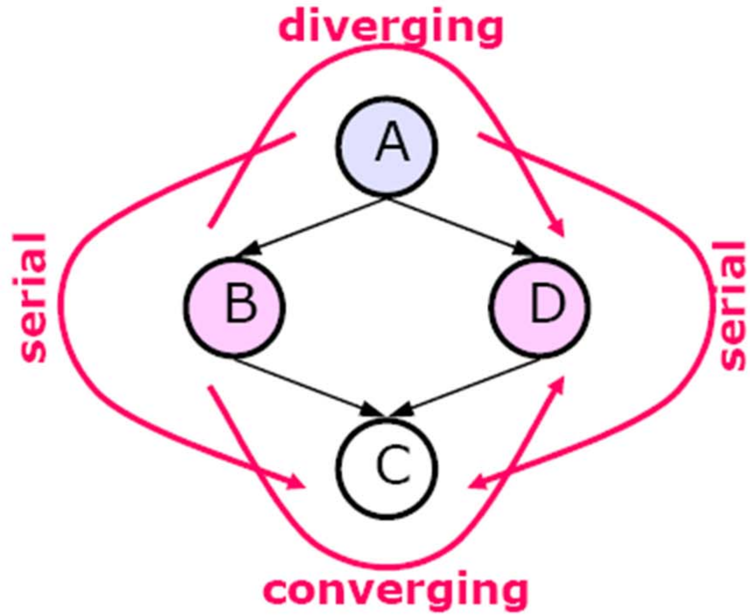
D-Separation Detail



- 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)

D-Separation Detail

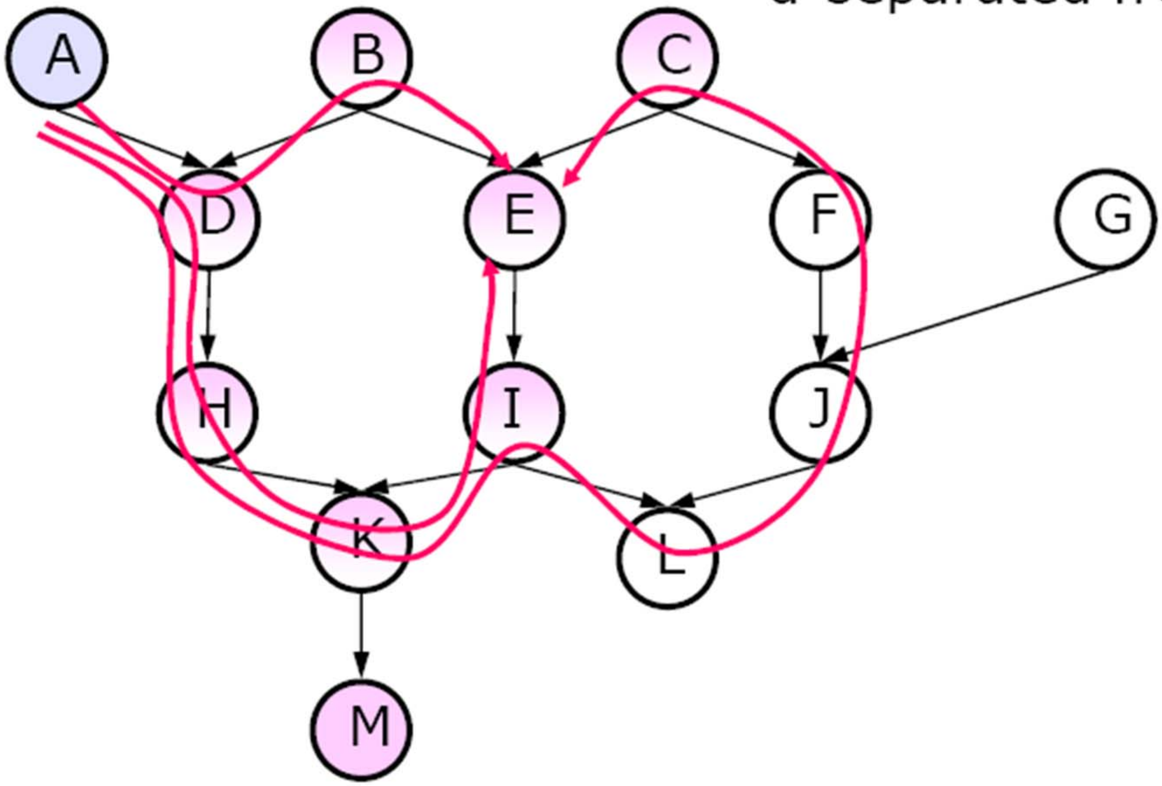


- A-B-C: serial, blocked when B is known, connected otherwise
- A-D-C: serial, blocked when D is known, connected otherwise
- B-A-D: diverging, blocked when A is known, connected otherwise
- B-C-D: converging, blocked when C has no evidence, connected otherwise

- 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)

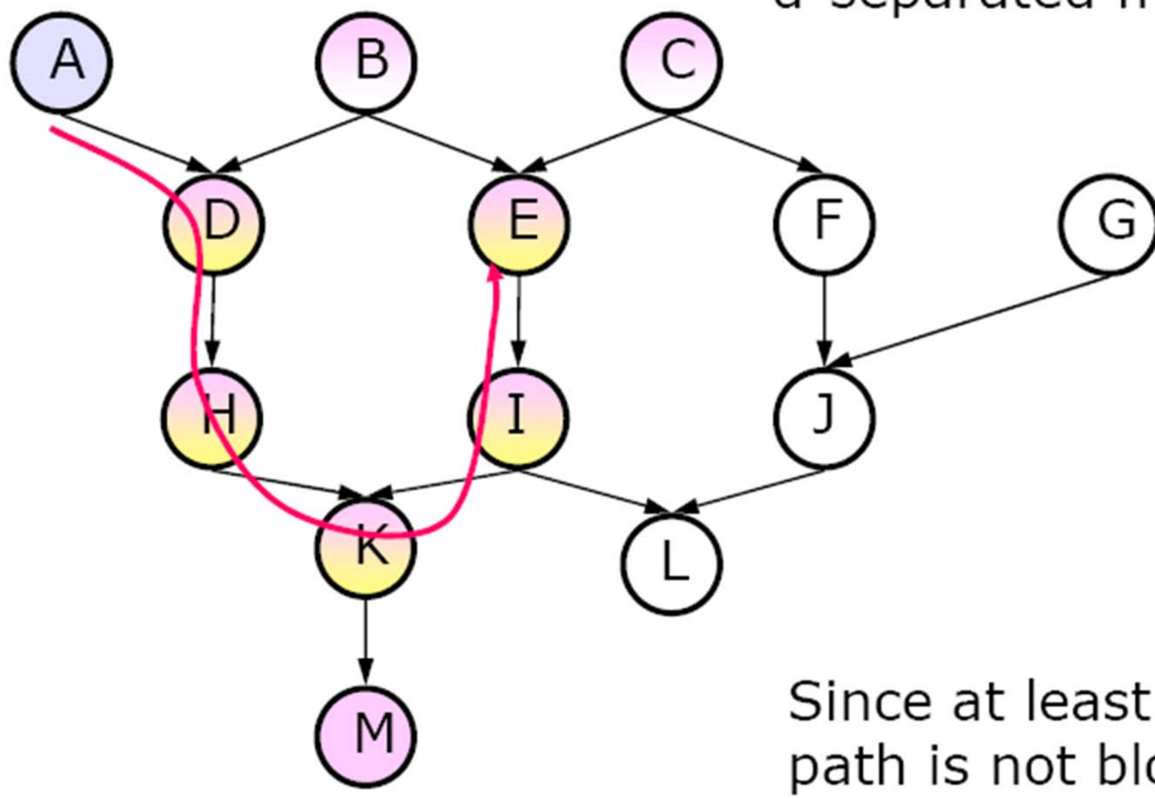
D-Separation Example

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



D-Separation Example

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



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

Bayesian (Belief) Networks

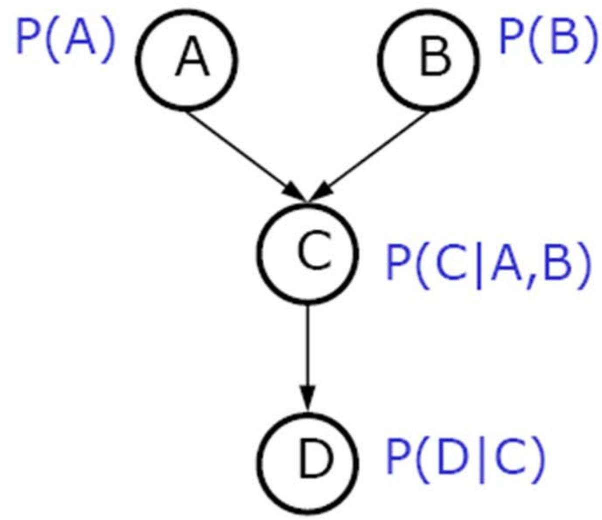
- Set of variables, each has a finite set of values
- Set of directed arcs between them forming acyclic graph
- Every node A, with parents $B_1, \ldots, B_n$, has $P(A \mid B_1, \ldots, B_n)$ specified

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

Chain Rule

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

$P(ABCD) = P(A{=}\text{true}, B{=}\text{true}, C{=}\text{true}, D{=}\text{true})$



$P(ABCD) =$

$P(D|ABC)P(ABC) =$

$P(D|C) \quad P(ABC) =$

$P(D|C) \quad P(C|AB)\, P(AB) =$

$P(D|C) \quad P(C|AB)\, P(A)P(B)$

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