# Artificial Intelligence

## CSC 665

tyler dae devlin

# PGMs III

*4.2.2024*

- **Search:** make decisions by looking ahead
- **Logic:** deduce new facts from existing facts
- **Constraints:** find a way to satisfy a given specification
- **Probability:** reason quantitatively about uncertainty
- **Learning:** make future predictions from past observations

# Independence helps

- Independence reduces the amount of data needed to specify the joint distribution
- Factored representation $P(X, Y) = P(X)P(Y)$
- 36 → 12 numbers in the dice example
- In general: exponential → linear in the number of variables
- **Problem:** independence is rare in practice

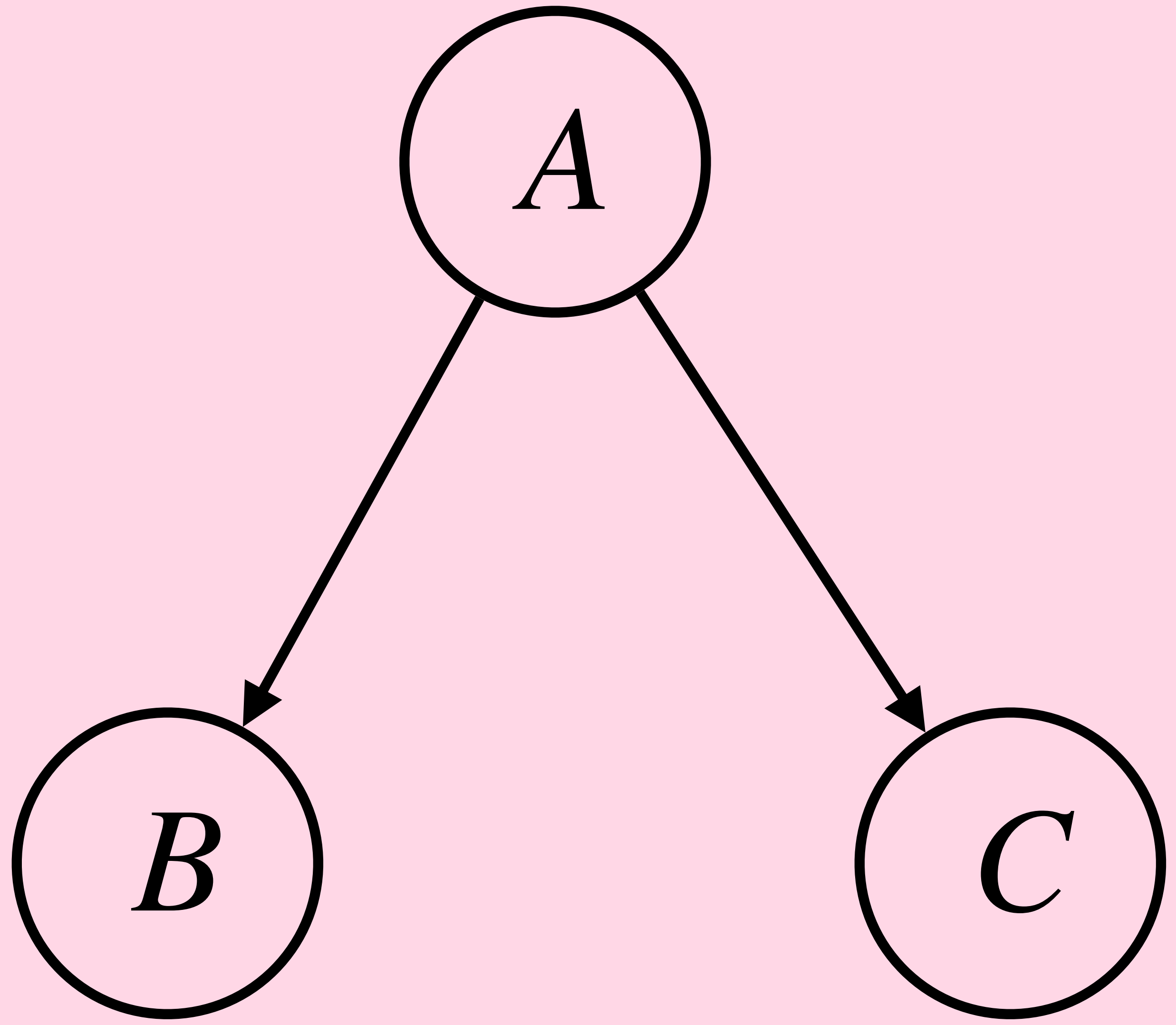
| *X* | *Y* | *P(X,Y)* |
|---|---|---|
| 1 | 1 | 1/36 |
| 1 | 2 | 1/36 |
| … | … | … |
| 6 | 6 | 1/36 |

| *X* | *P(X)* |
|---|---|
| 1 | 1/6 |
| 2 | 1/6 |
| … | … |
| 6 | 1/6 |

| *Y* | *P(Y)* |
|---|---|
| 1 | 1/6 |
| 2 | 1/6 |
| … | … |
| 6 | 1/6 |

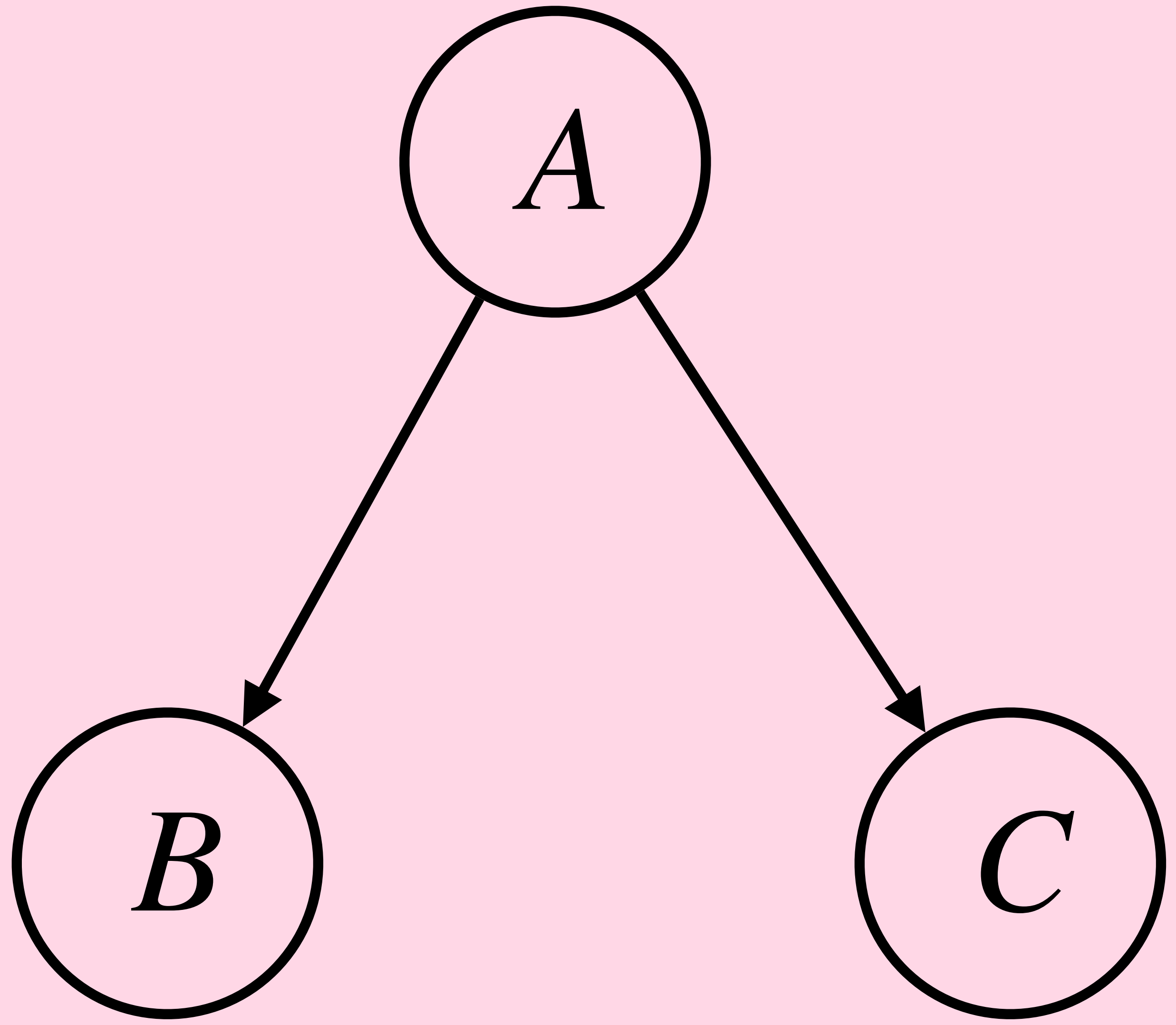
# Example: alarm reporting

- Your house has an alarm system
- If the alarm goes off, your two neighbors Bob and Charlie will call you if they hear it
- Binary random variables $A$, $B$, $C$ for alarm goes off, Bob calls, Charlie calls
- **Question:** are $B$ and $C$ independent?
- No! If Bob calls, then that means the alarm likely went off, so Charlie is likely to call too
- $P(C = 1 \mid B = 1) > P(C = 1)$



# Example: alarm reporting

- $B$ and $C$ are not independent: $P(C \mid B) \neq P(C)$
- But they are *conditionally independent* given $A$
- If the alarm goes off, Charlie's ability to hear it is not affected by Bob
- $P(C \mid A, B) = P(C \mid A)$
- This allows for a factored joint distribution
- $P(B, C \mid A) = P(B \mid A)P(C \mid A)$

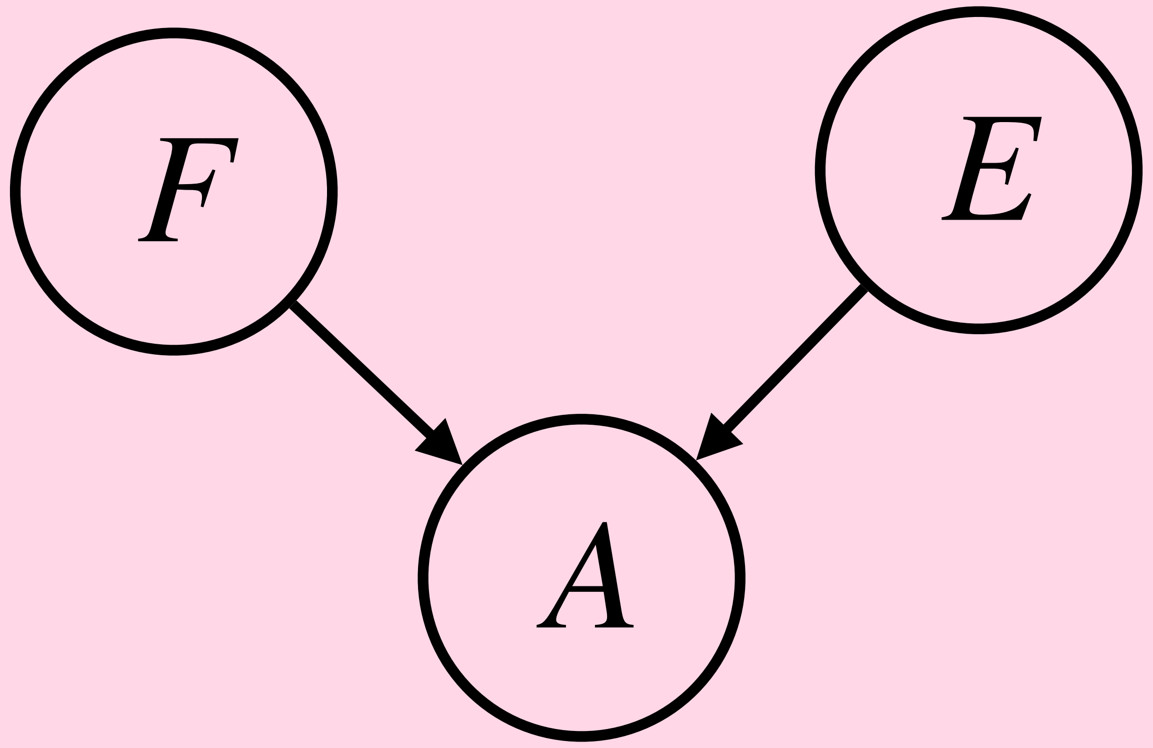


# Conditional independence

- $X$ and $Y$ are conditionally independent given $Z$ if $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$
- Conditional independence is often easier to find in practice than unconditional independence
- **Important:** independence is different from conditional independence. Can have one without the other.

# Example: alarm triggering

- Assume fires and earthquakes are independent rare events
- Either one will set off the alarm
- Joint distribution
  $P(F, E, A) = P(F)P(E)P(A \mid F, E)$
- Compute:
  - $P(E = 1 \mid A = 1)$
  - $P(E = 1 \mid A = 1, F = 1)$



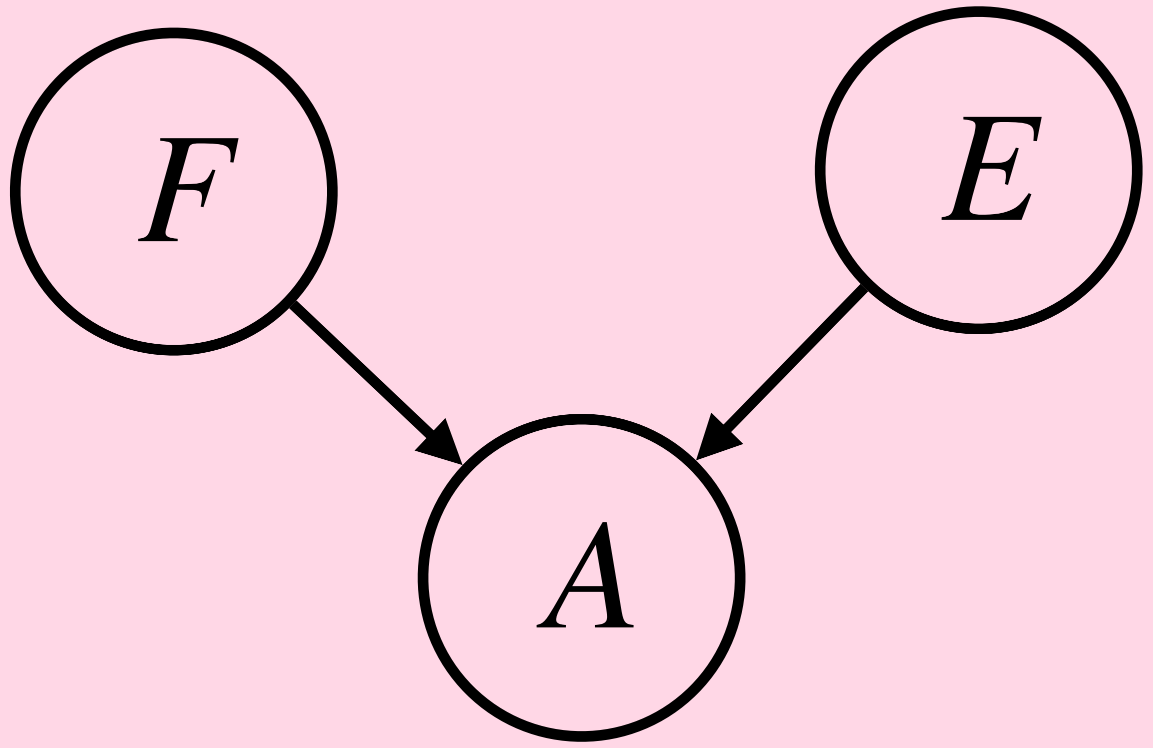

| $P(F=1)$ |
|---|
| $\epsilon$ |

| $P(E=1)$ |
|---|
| $\epsilon$ |

| $f$ | $e$ | $P(A=1 \mid F=f, E=e)$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

# Example: alarm triggering

- Assume fires and earthquakes are independent rare events
- Either one will set off the alarm
- Joint distribution
  $P(F, E, A) = P(F)P(E)P(A \mid F, E)$
- Compute:
  - $P(E = 1 \mid A = 1) = \frac{\epsilon(1-\epsilon) + \epsilon^2}{\epsilon(1-\epsilon) + \epsilon^2 + (1-\epsilon)\epsilon} = \frac{1}{2-\epsilon}$
  - $P(E = 1 \mid A = 1, F = 1) = \frac{\epsilon^2}{\epsilon^2 + (1-\epsilon)\epsilon} = \epsilon$



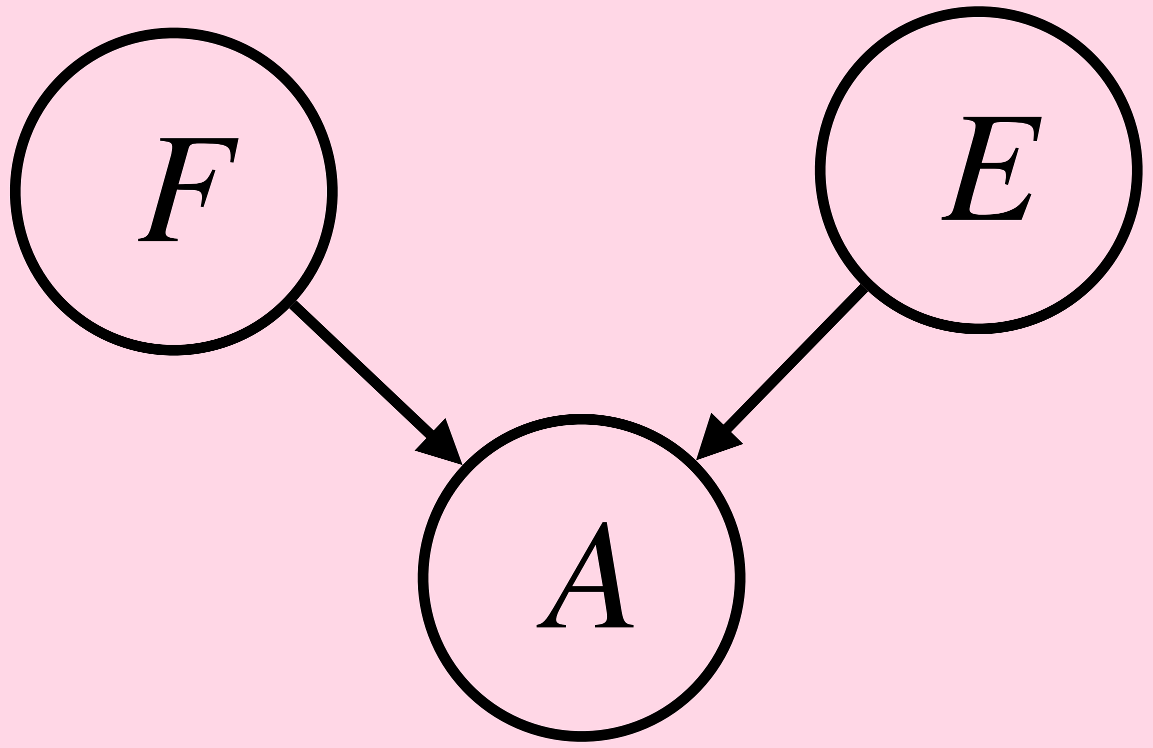

| **P(F = 1)** |
|---|
| ϵ |

| **P(E = 1)** |
|---|
| ϵ |

| **f** | **e** | **P(A = 1 \| F = f, E = e)** |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

# Example: alarm triggering

- $P(E = 1 \mid A = 1, F = 1) < P(E = 1 \mid A = 1)$
- **Conclusion:** If your alarm goes off, knowing there was a fire decreases the chance that there was an earthquake
- The fire has “explained away” the alarm
- Not a causal statement: fires do not protect against earthquakes!
- $F$ and $E$ are independent (unconditionally)
- But $F$ and $E$ are conditionally *dependent* given $A$!



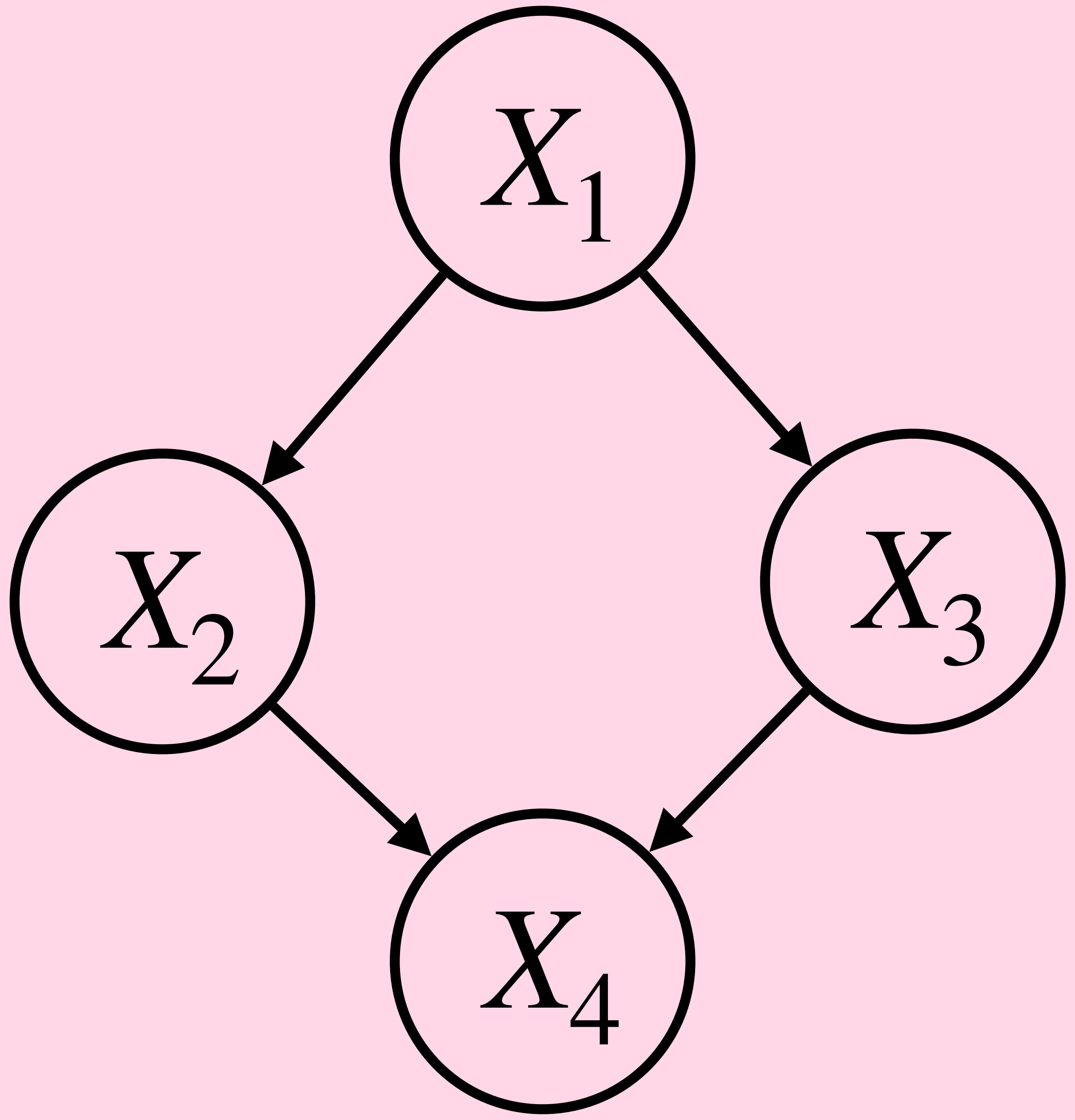

| $P(F = 1)$ |
|---|
| $\epsilon$ |

| $P(E = 1)$ |
|---|
| $\epsilon$ |

| $f$ | $e$ | $P(A = 1 \mid F = f, E = e)$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

# Bayesian networks

- Let $X = (X_1, \ldots, X_n)$ be random variables
- A Bayesian network is a directed acyclic graph (DAG) where each node is a random variable
- The Bayesian network specifies a joint distribution over $X$ as a product of local conditional distributions, one for each node
- $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{parents}(X_i))$$



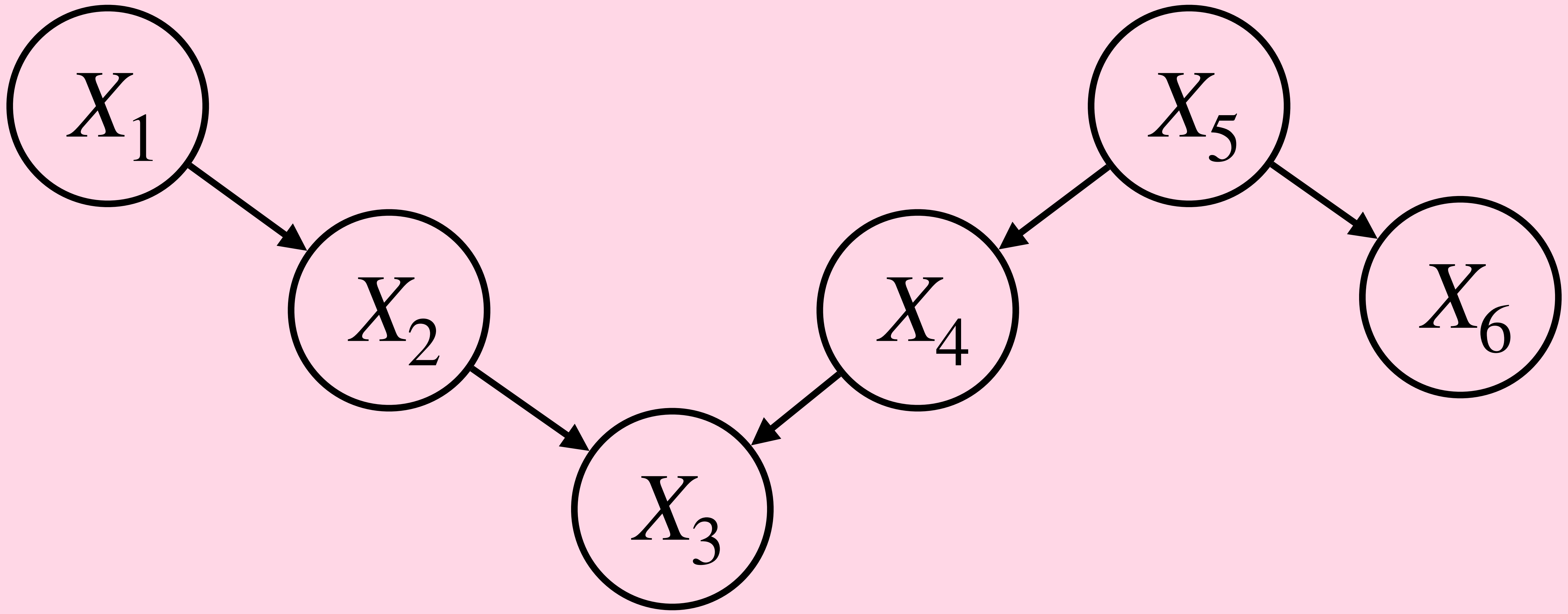

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4 \mid X_2, X_3)$$

# d-separation

- A Bayesian network lets us read off the conditional independence relationships between any pair of variables
- d-separation in the graph $\iff$ conditional independence in the joint distribution

# d-separation rules

- **Rule 1:** unconditional d-separation
- $X$ and $Y$ are d-separated if there is no unblocked path between them
- A path is blocked if it contains two arrows colliding head-to-head in a v-structure

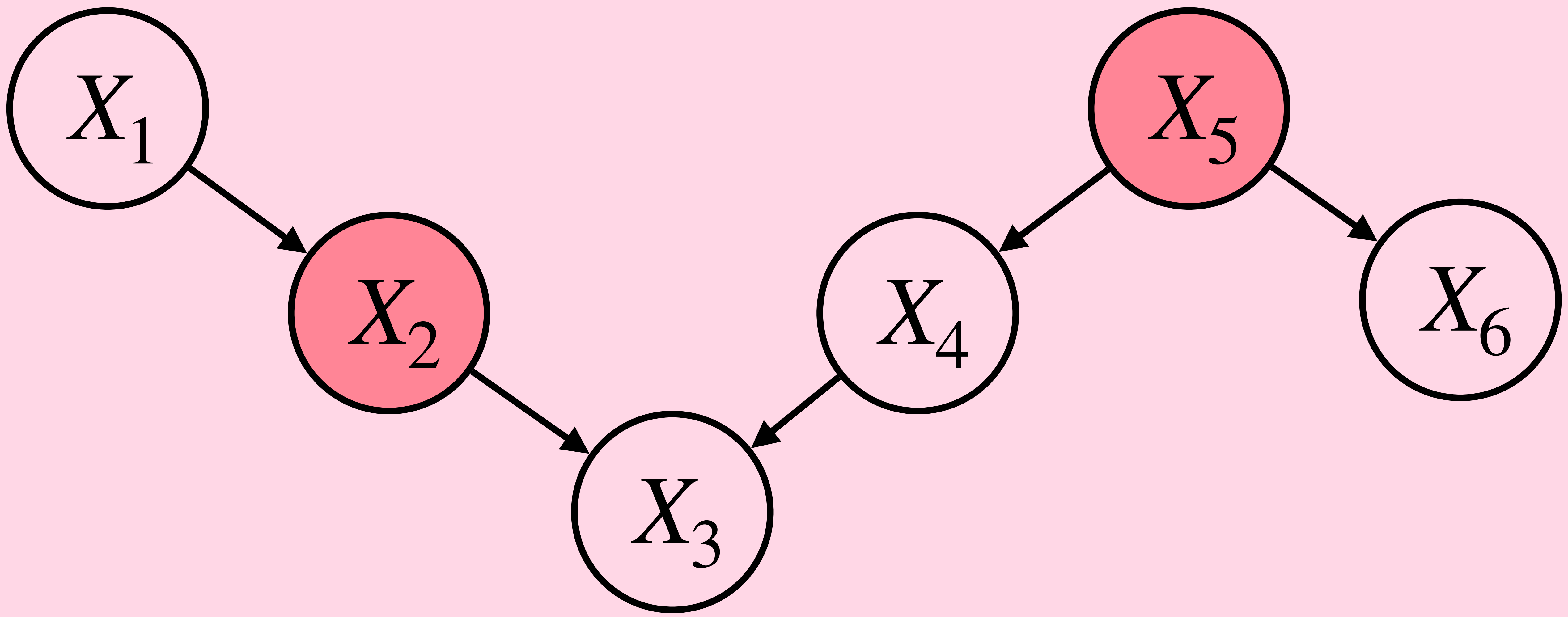



**d-separated:** $X_2$ and $X_4$, $X_1$ and $X_6$

**not d-separated:** $X_1$ and $X_3$, $X_4$ and $X_5$, $X_3$ and $X_6$

# d-separation rules

- **Rule 2:** blocking by conditioning
- An unblocked path becomes blocked if one of the nodes in the path is observed (shaded)

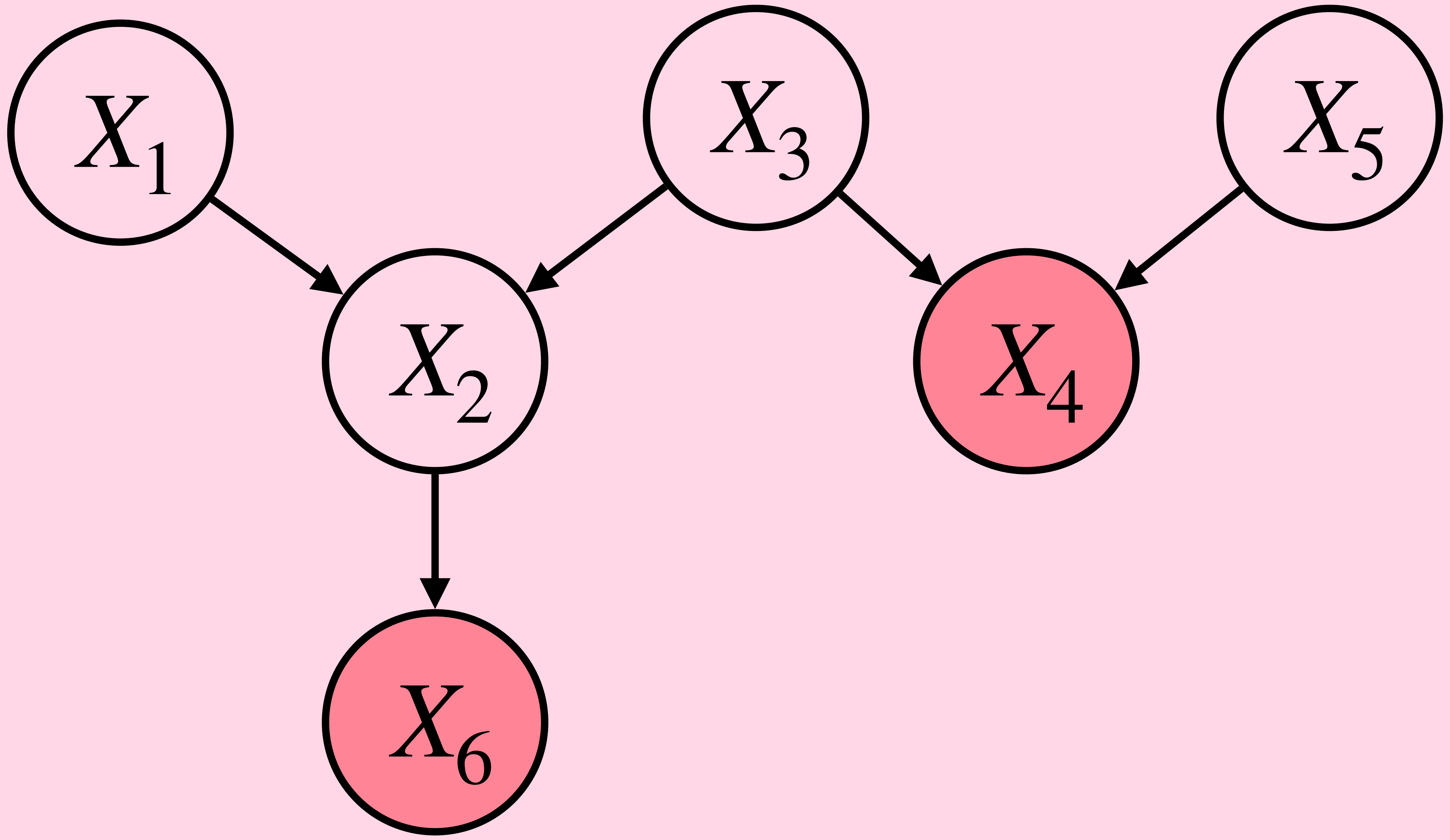



**d-separated:** $X_1$ and $X_3$, $X_3$ and $X_6$, $X_1$ and $X_6$

**not d-separated:** $X_3$ and $X_4$

# d-separation rules

- **Rule 3:** activated v-structures
- If the node at the center of a v-structure or one of its descendants is observed, then the v-structure is no longer blocking

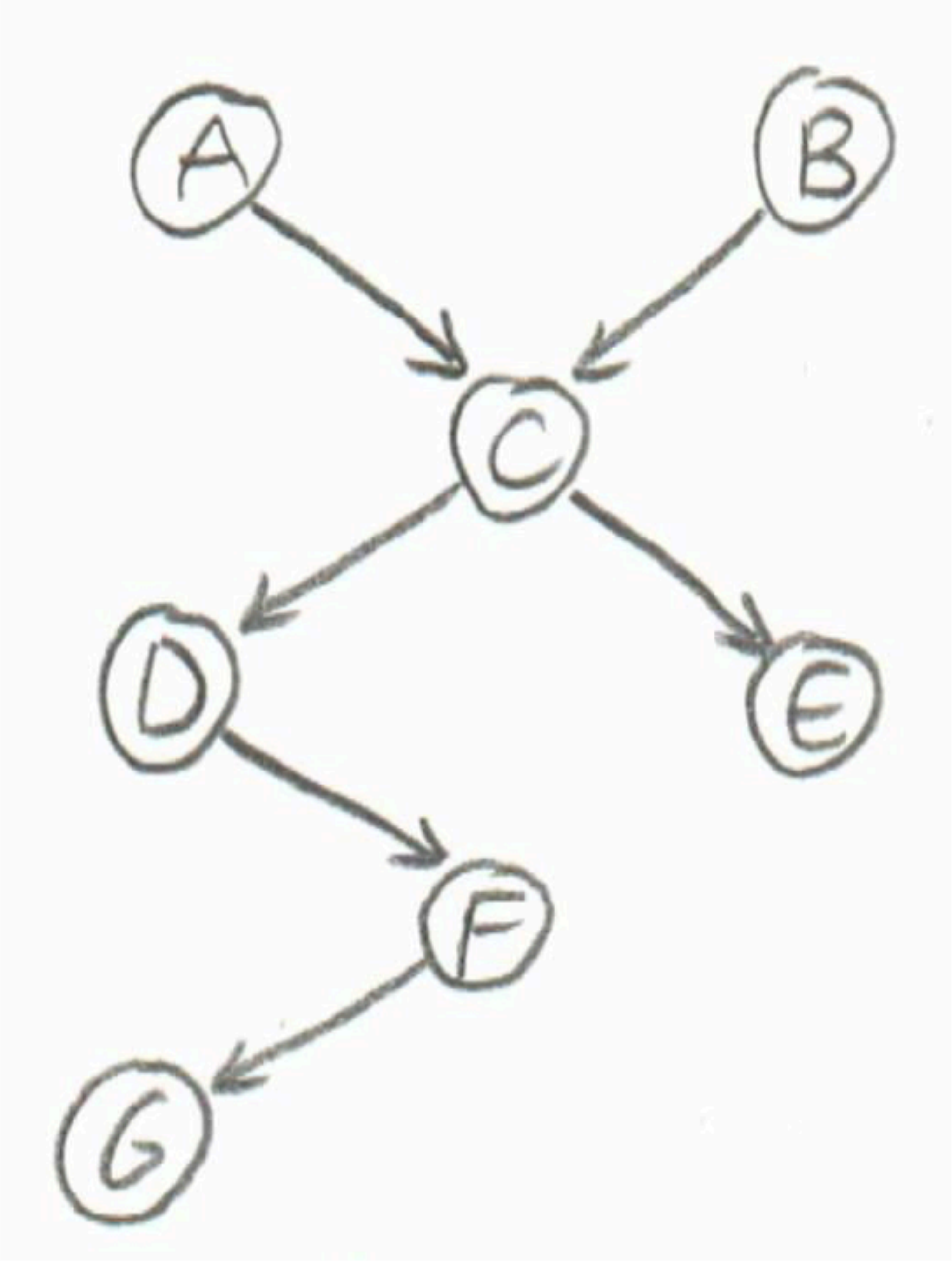



**d-separated:** nothing

**not d-separated:** everything

# d-separation examples

- A and B, given D and F
- A and B
- A and B, given C
- D and E, given C
- D and E
- D and E, given A and B



# d-separation examples

- A and B, given D and F
- **A and B**
- A and B, given C
- **D and E, given C**
- D and E
- D and E, given A and B

(**bold** indicates d-separation)