6.034 FALL 2020

RECITATION 5: BAYES pt. 2

- [ ] PROBABILITY RECAP
- [ ] BAYES NETS
  - [ ] D-SEPARATION
  - [ ] # OF PARAMETERS
  - [ ] NAIVE BAYES CLASSIFICATION
  - [ ] MODEL SELECTION

Applications of Bayes Nets

PROBABILITY: Types of Probability

$\Omega = \{A, B, C\}$

- Joint
  - $P(A=a, B=b, C=c)$
- Marginal
  - $P(A=a) = \sum_b \sum_c P(A=a, B=b, C=c)$
- Conditional
  - $P(A=a \mid B=b, C=c) = \frac{P(A=a, B=b, C=c)}{P(B=b, C=c)}$ $\left(\frac{\text{joint}}{\text{marginal}}\right)$

PROBABILITY: Chain & Bayes Rules

$\Omega = \{A, B\}$

- ☐ Chain rule
  - ☐ $P(A=a, B=b) = P(A=a \mid B=b)\,P(B=b)$
    $= P(B=b \mid A=a)\,P(A=a)$
  - ☐ $P(A=a \mid B=b)\,P(B=b) = P(B=b \mid A=a)\,P(A=a)$
- ☐ Bayes Rule
  - ☐ $P(A=a \mid B=b) = \dfrac{P(B=b \mid A=a)\,P(A=a)}{P(B=b)}$

PROBABILITY: Independence

Marginal Independence

$\square$ $A \perp\!\!\!\perp B$ iff $P(A|B) = P(A)$

$\Rightarrow$ $\square$ $P(A,B) = P(A|B)P(B) = P(A)P(B)$

Conditional Independence

$\square$ $(A \perp\!\!\!\perp B | C)$ iff $P(A|B,C) = P(A|C)$

$\Rightarrow$ $\square$ $P(A,B|C) = P(A|C)P(B|C)$

PROBABILITY: # of parameters

$\Omega = \{A, B\}$ (Boolean)

if $A \not\perp\!\!\!\perp B$:

$$\#\text{ params} = \left[\prod_{i}^{n} \#\text{vals of } i\right] - 1$$

if A & B Boolean: $2^n - 1$

if $A \perp\!\!\!\perp B$

$$\#\text{ params} = \sum^{n} \left[(\#\text{vals of } i) - 1\right]$$

if A & B Boolean: n

BAYES NETS: Bayes Net assumption

<u>Def:</u> Bayes Net

- ☐ directed, acyclic graph
- ☐ represents the independence assumption

<u>Assumption</u>

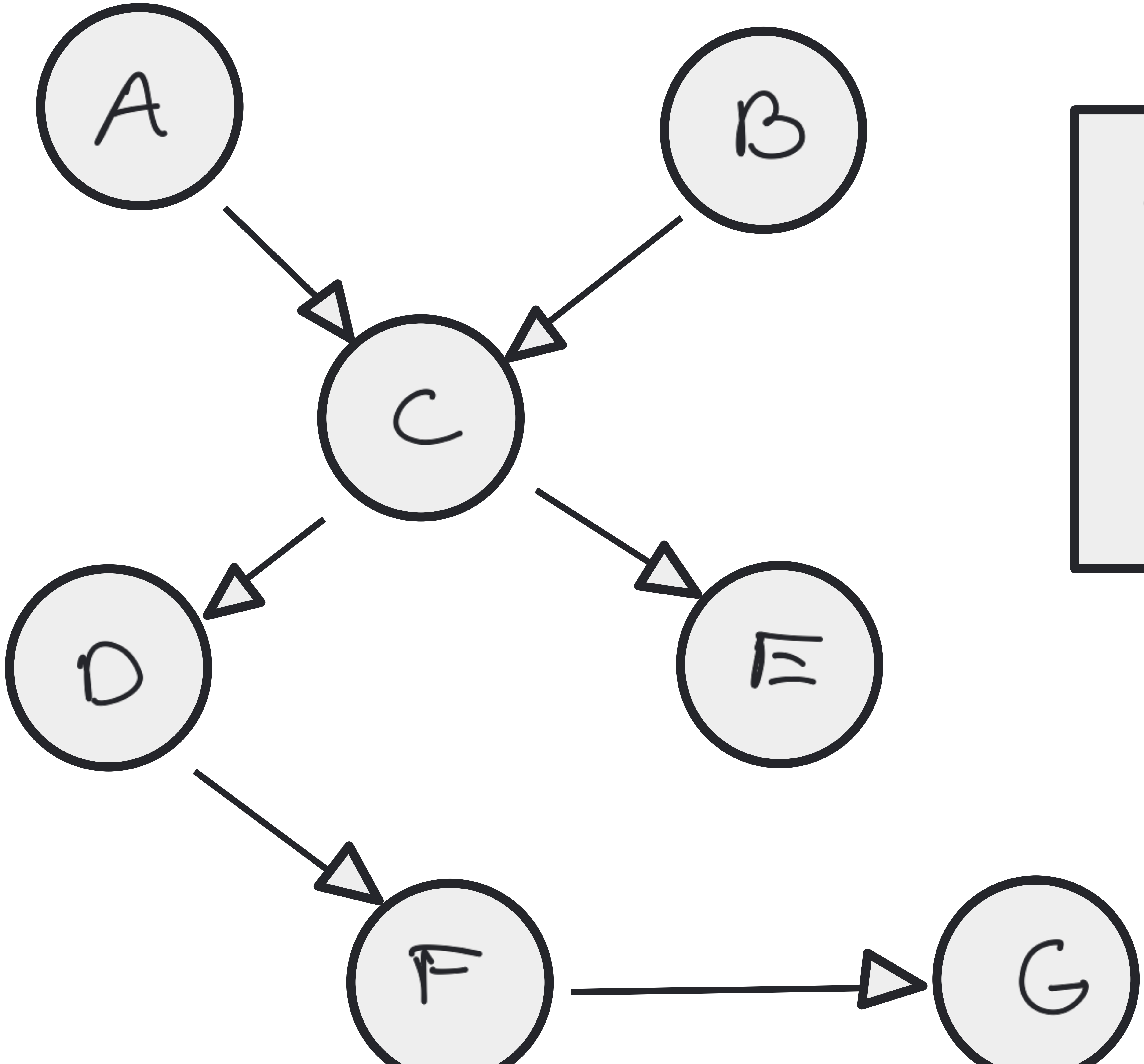
★ Each variable in the Bayes net is conditionally independent of its nondescendants given its parents

BAYES NETS: d-separation

1) Draw ancestral graph
2) Link the parents
3) Disorient the graph
4) Delete givens & edges
5) Read:
    - if variables connected by a path: $A \not\perp\!\!\!\perp B$*
    - if not: $A \perp\!\!\!\perp B$

(* A & B are <u>structurally</u> not independent)

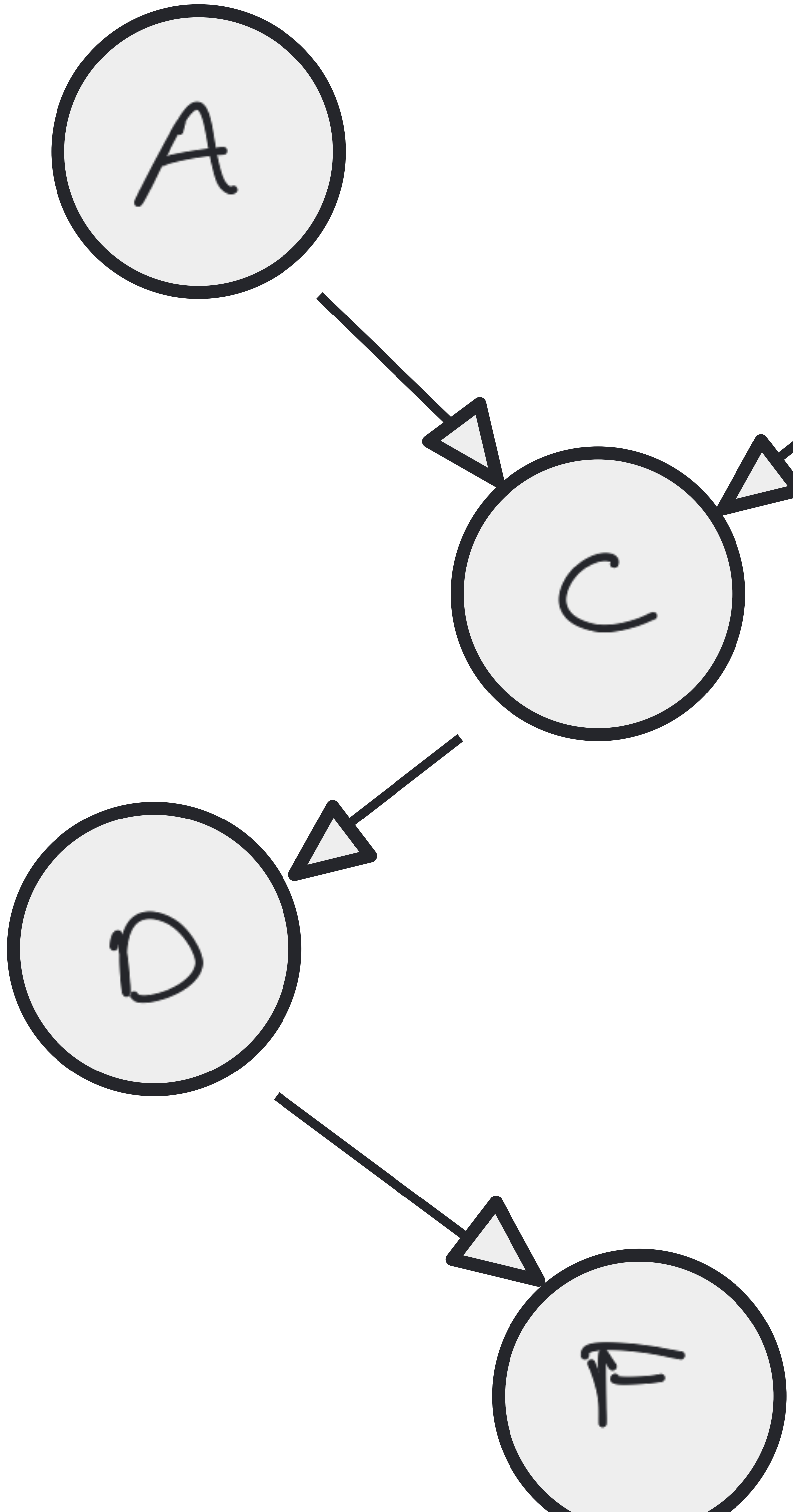
BAYES NETS: d-separation



Want to know:

- ☐ $A \perp\!\!\!\perp B \mid D, F$?
- ☐ $P(A \mid B, D, F) \overset{?}{=} P(A \mid D, F)$?

BAYES NETS: d-separation

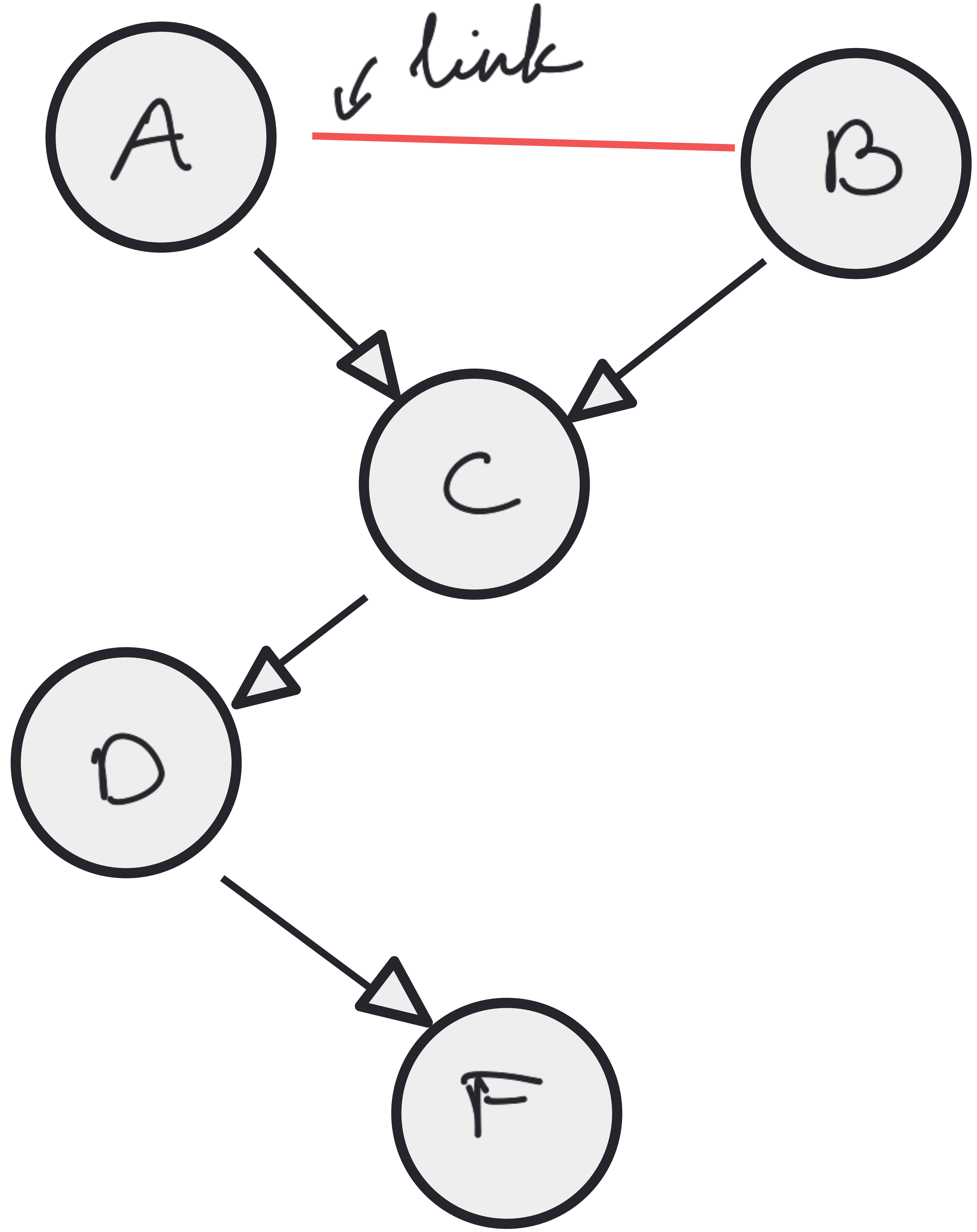


Want to know:

- ☐ $A \perp\!\!\!\perp B \mid D, F$?
- ☐ $P(A \mid B, D, F) \stackrel{?}{=} P(A \mid D, F)$?

① Draw an ancestral graph

BAYES NETS: d-separation

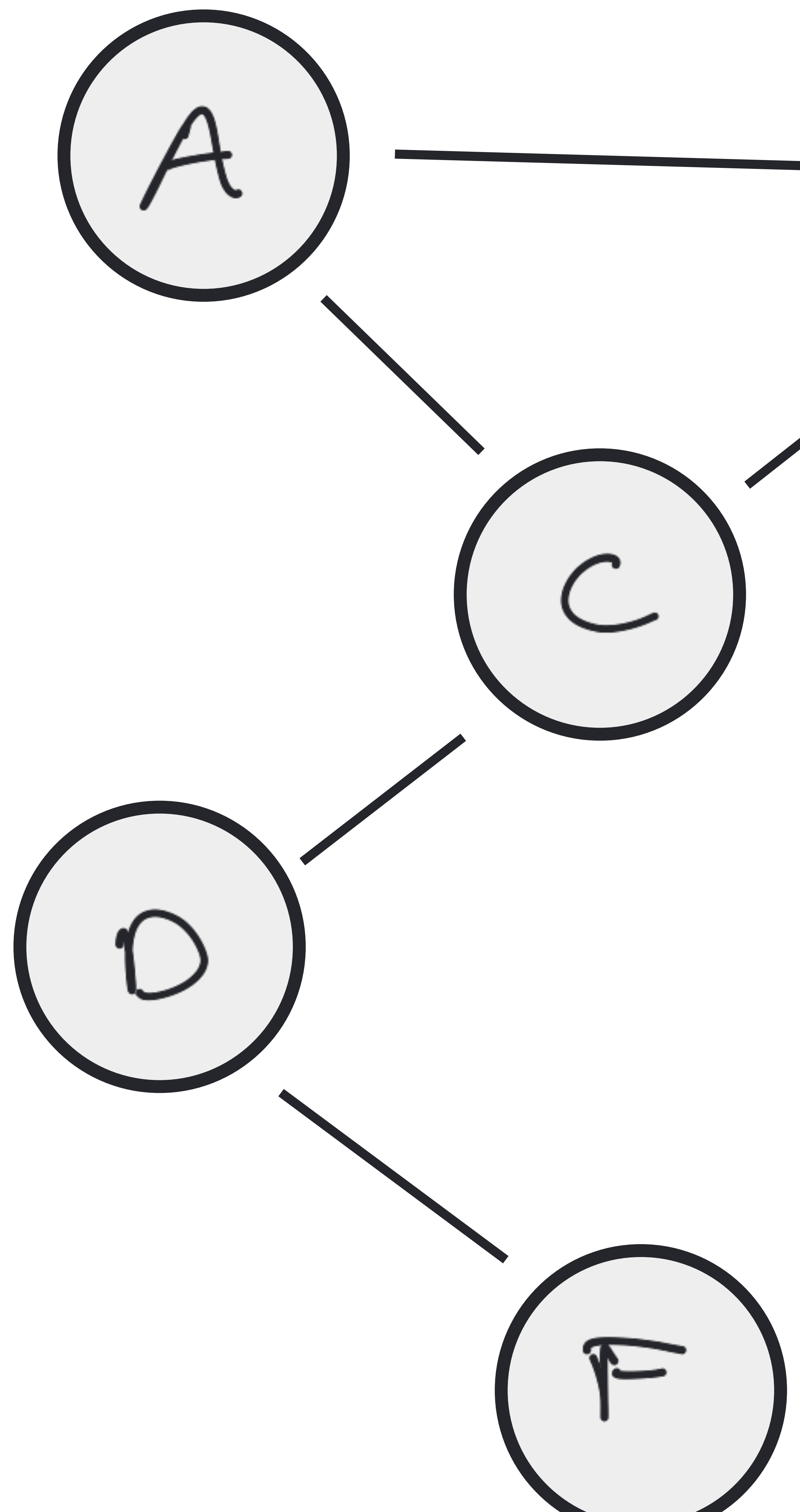


Want to know:

- ☐ $A \perp\!\!\!\perp B \mid D, F$?
- ☐ $P(A \mid B, D, F) \stackrel{?}{=} P(A \mid D, F)$?

② Link the parents

BAYES NETS: d-separation

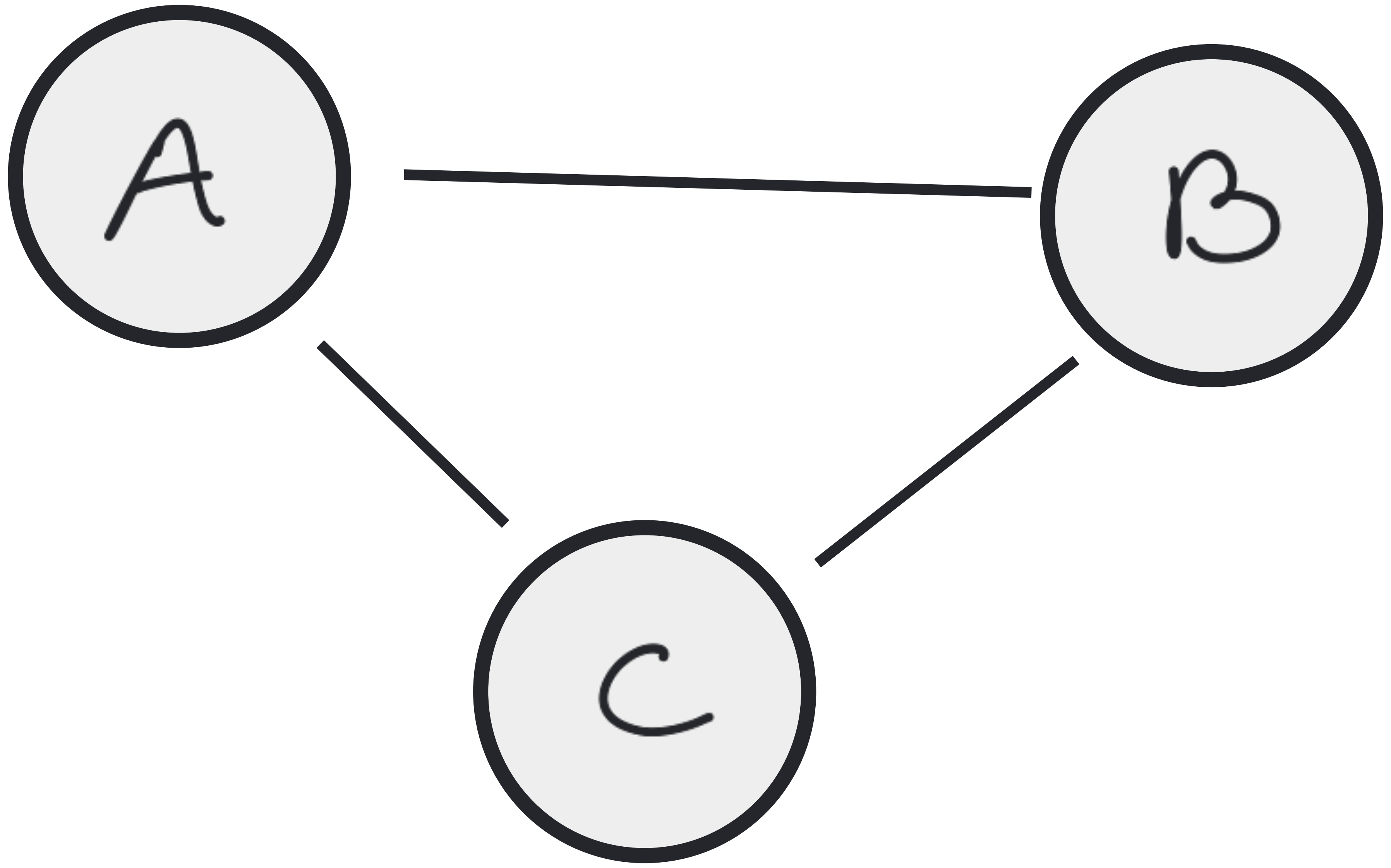


Want to know:

- ☐ $A \perp\!\!\!\perp B \mid D, F$?
- ☐ $P(A \mid B, D, F) \stackrel{?}{=} P(A \mid D, F)$?

③ Disorientation

BAYES NETS: d-separation

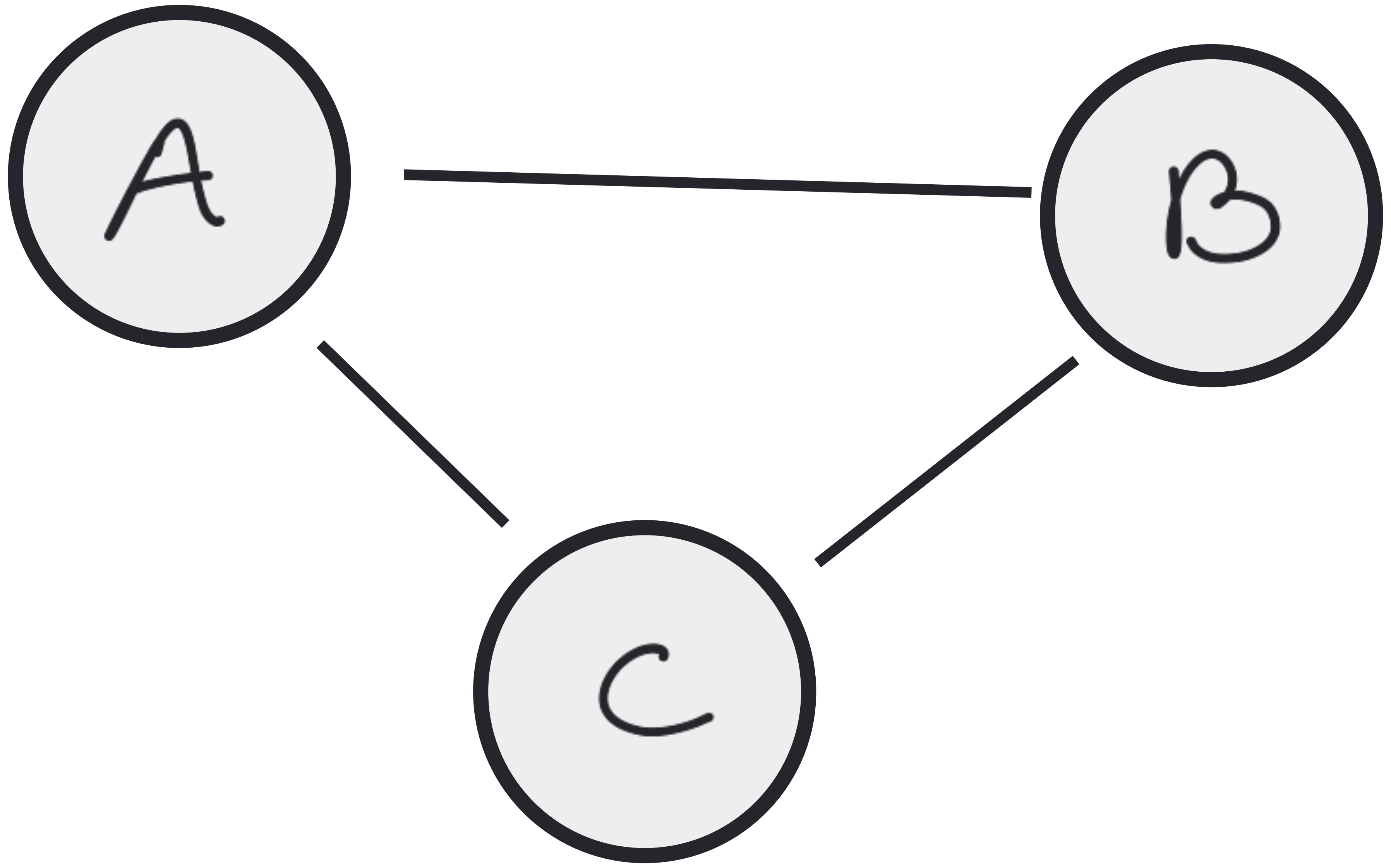


Want to know:

- ☐ $A \perp\!\!\!\perp B \mid D, F$?
- ☐ $P(A \mid B, D, F) \stackrel{?}{=} P(A \mid D, F)$?

④ Delete givens & their edges

BAYES NETS: d-separation



Want to know:

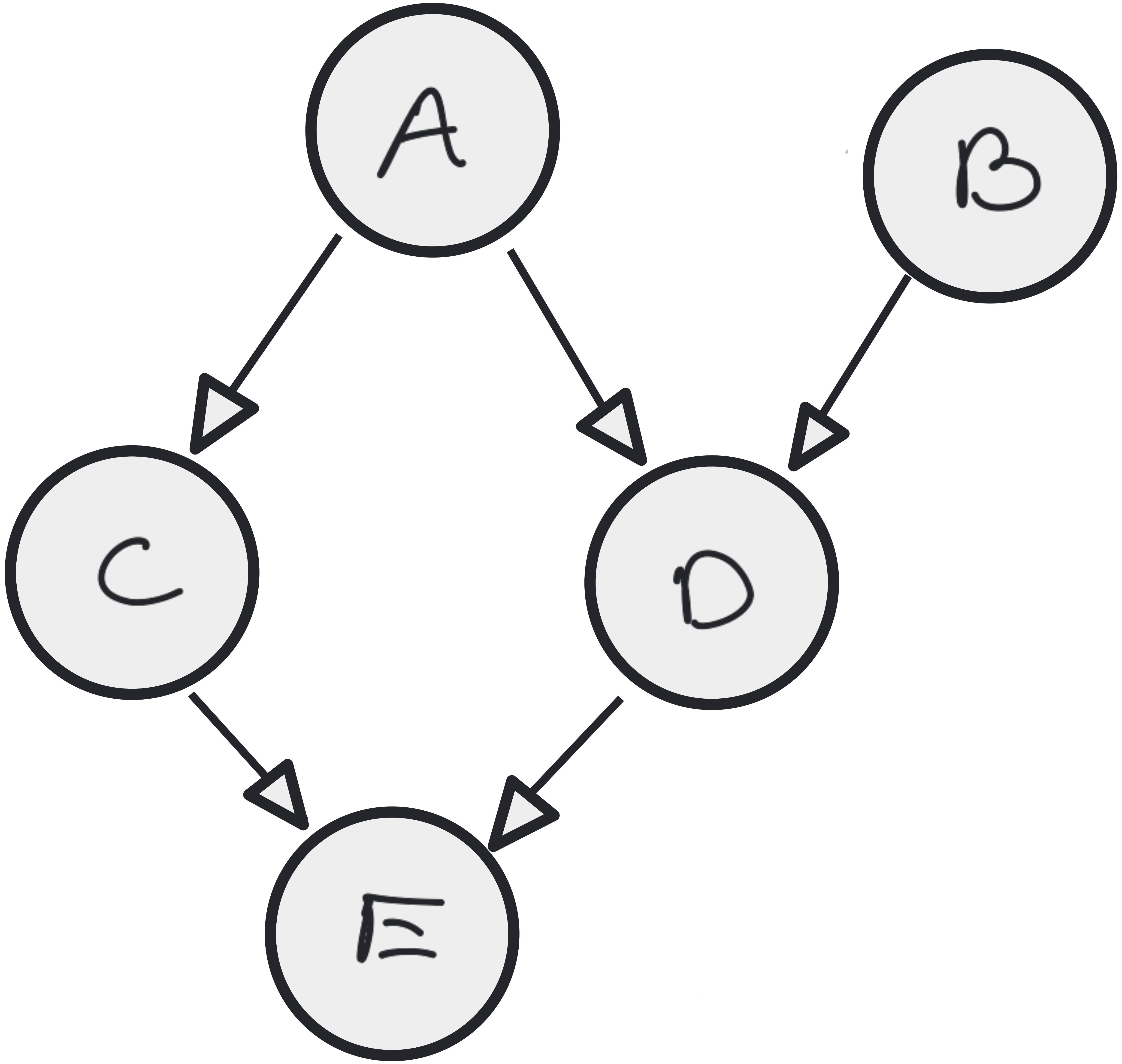
- ☐ $A \perp\!\!\!\perp B \mid D, F$?
- ☐ $P(A \mid B, D, F) \stackrel{?}{=} P(A \mid D, F)$?

⑤ Read the graph

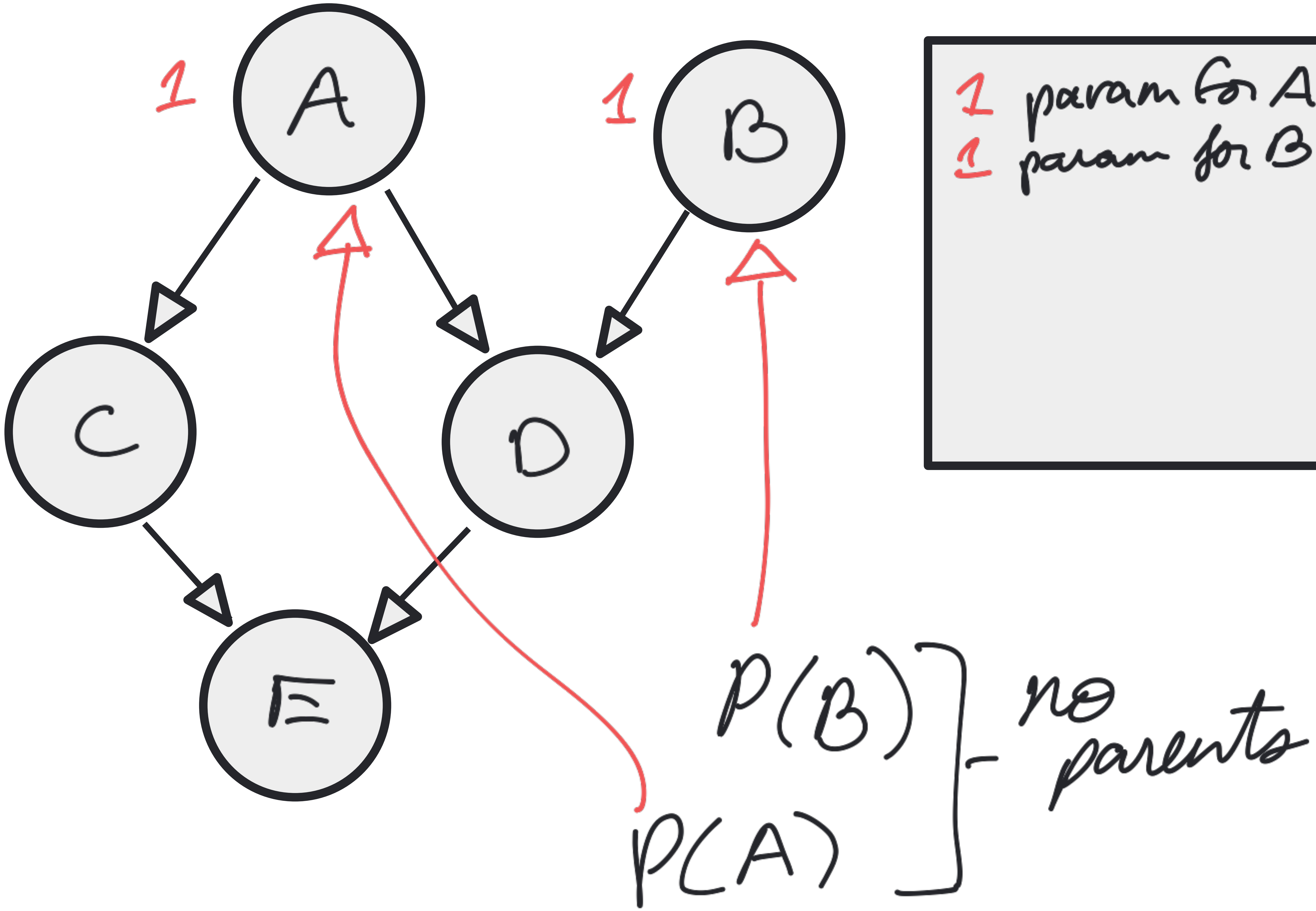
Conclusion: $A \not\perp\!\!\!\perp B \mid D, F$ *

* A & B are structurally $\not\perp\!\!\!\perp$, but mathematically may still be $\perp\!\!\!\perp$

BAYES NETS: # of Parameters

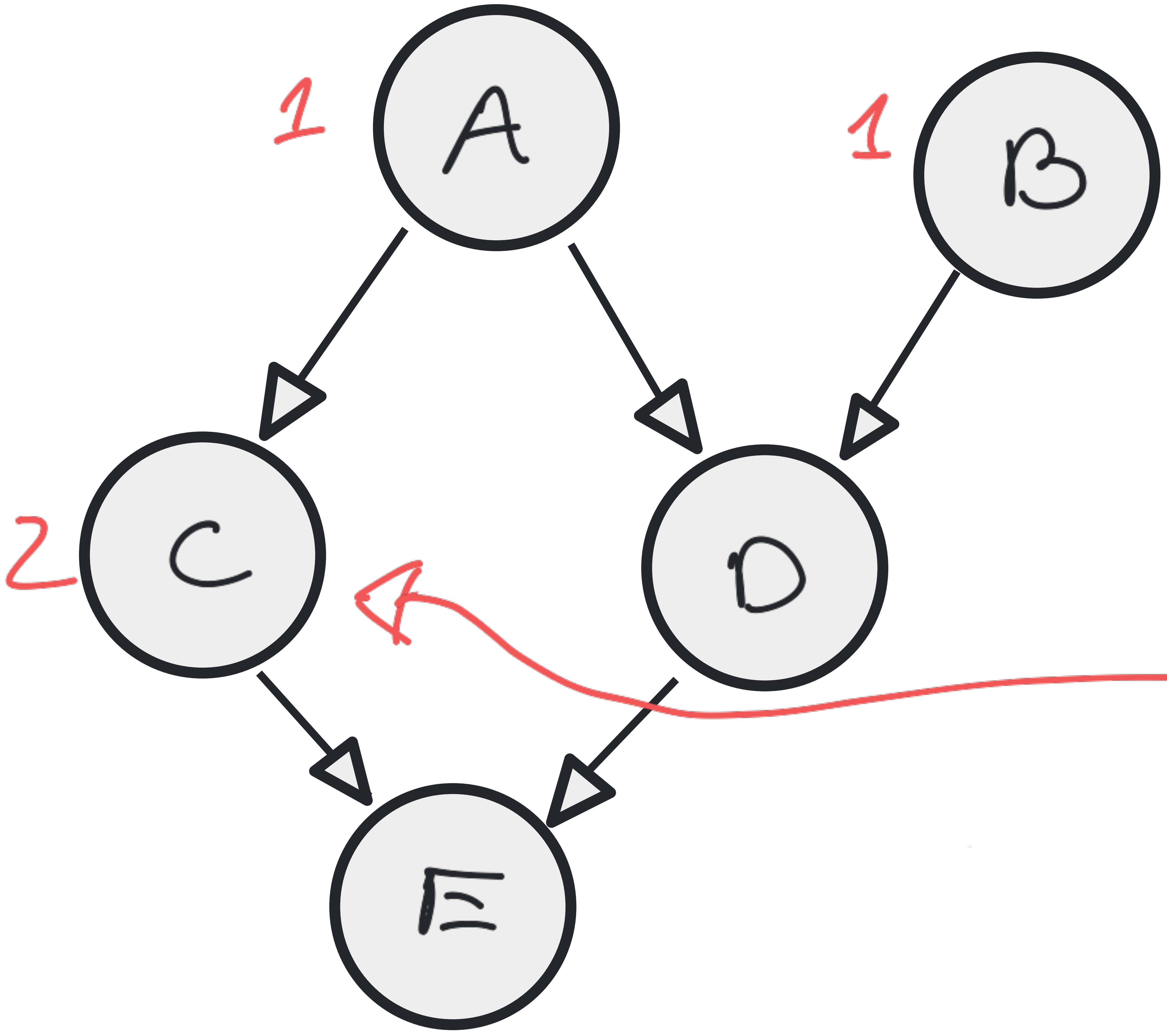


BAYES NETS: # of Parameters



1 param for A
1 param for B

BAYES NETS: # of Parameters



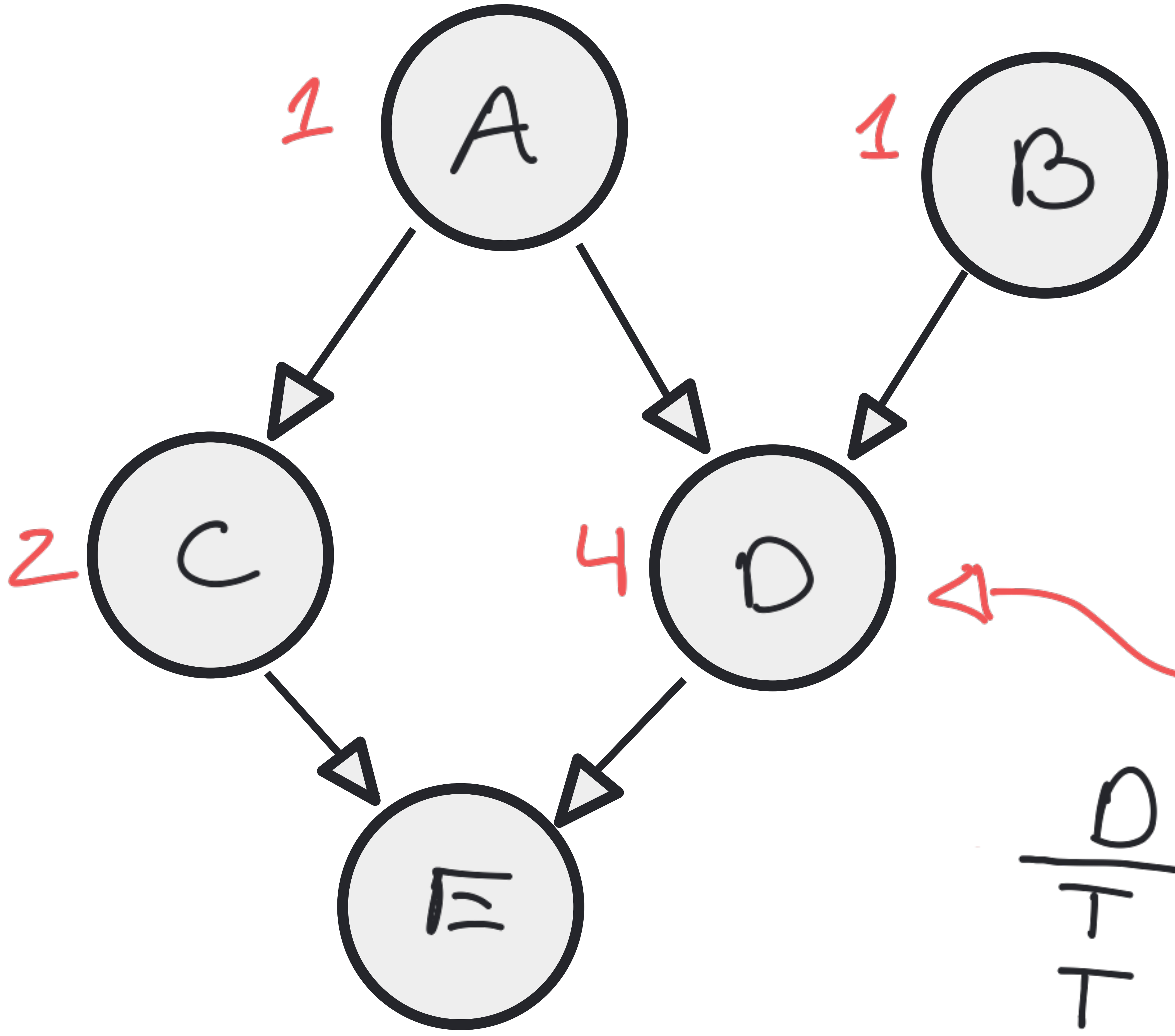
1 param for A]- no parents
1 param for B]
2 params for C

$P(C|A)$

C	A	$P(C|A)$
T	T	p_1
T	F	p_2
F	T	$1-p_1$
F	F	$1-p_2$

exhaustion principle

BAYES NETS: # of Parameters



1 param for A] - no parents
1 param for B]
2 params for C (1 parent)
4 params for D

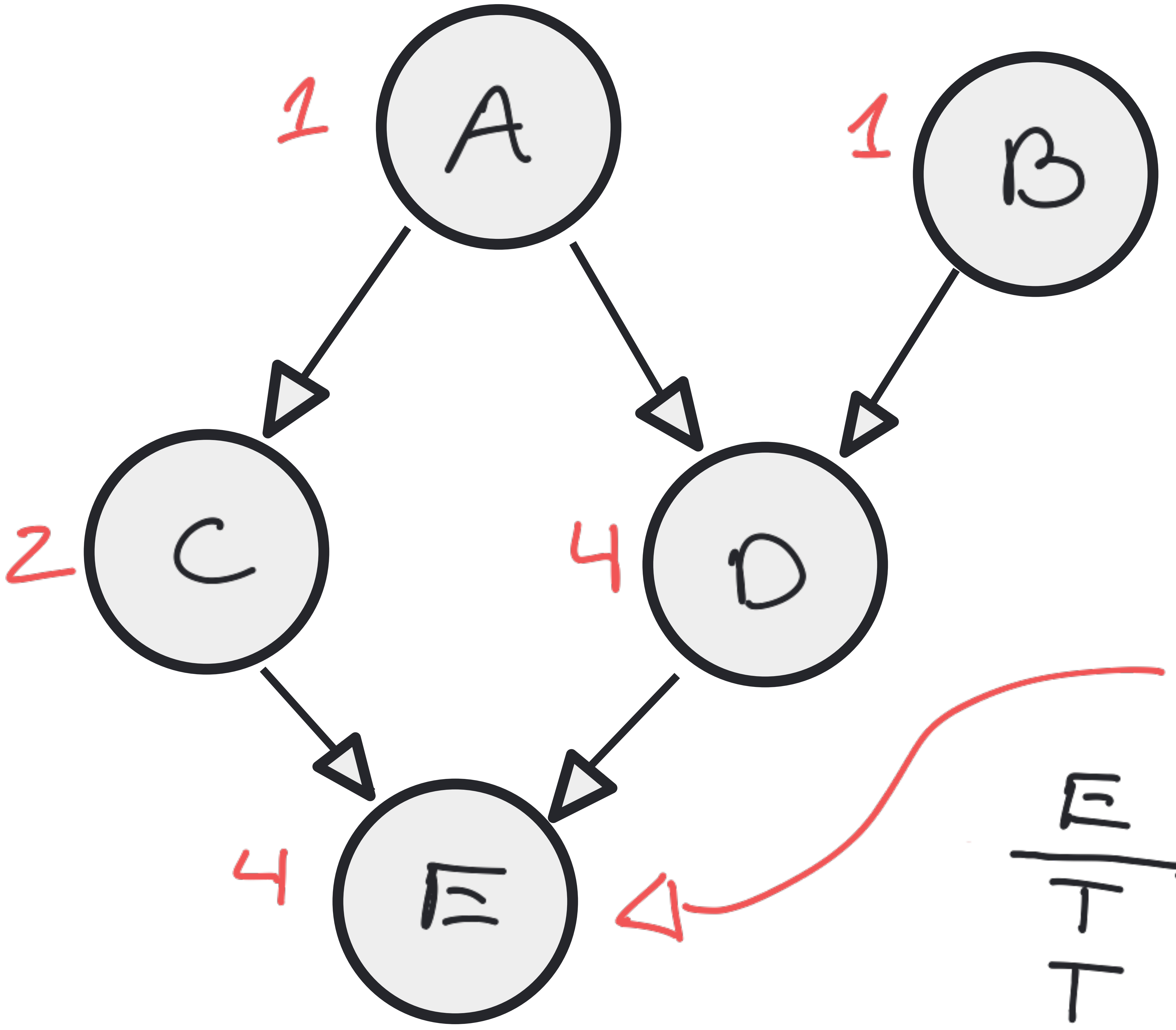
$P(D \mid A, B)$

D	A	B	$P(D \mid A,B)$
T	T	T	p_1
T	T	F	p_2
T	F	T	p_3
T	F	F	p_4
F	T	T	$1-p_1$
F	T	F	

4 parameters

exhaustion

BAYES NETS: # of Parameters



1 param for A]- no parents
1 param for B]
2 params for C (1 parent)
4 params for D (2 parents)
4 params for E

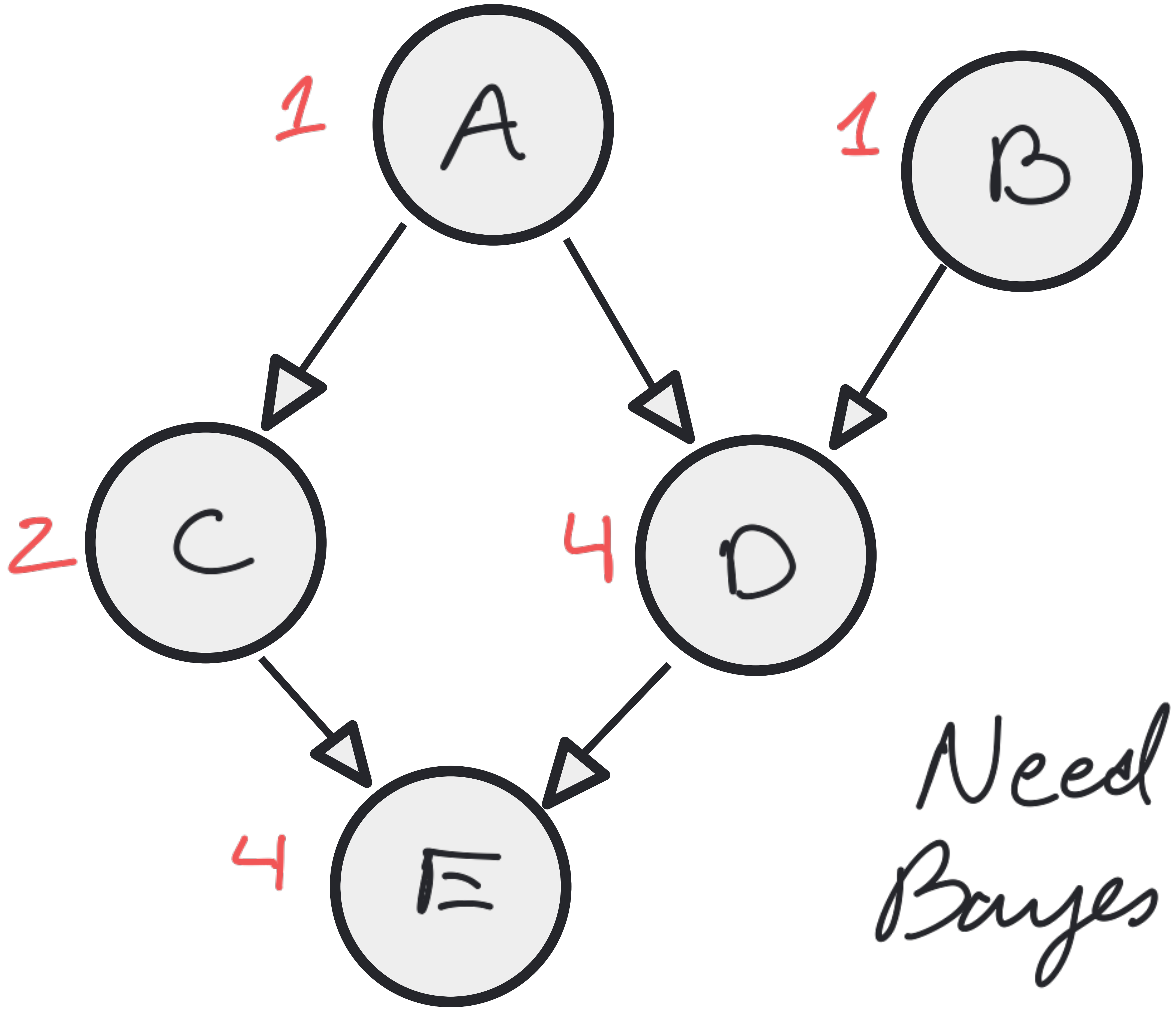
$P(E \mid C, D)$

E	C	D	$P(E \mid C, D)$
T	T	T	p_1
T	T	F	p_2
T	F	T	p_3
T	F	F	p_4
F	T	T	$1-p_1$
F	T	F	

4 parameters

exhaustion

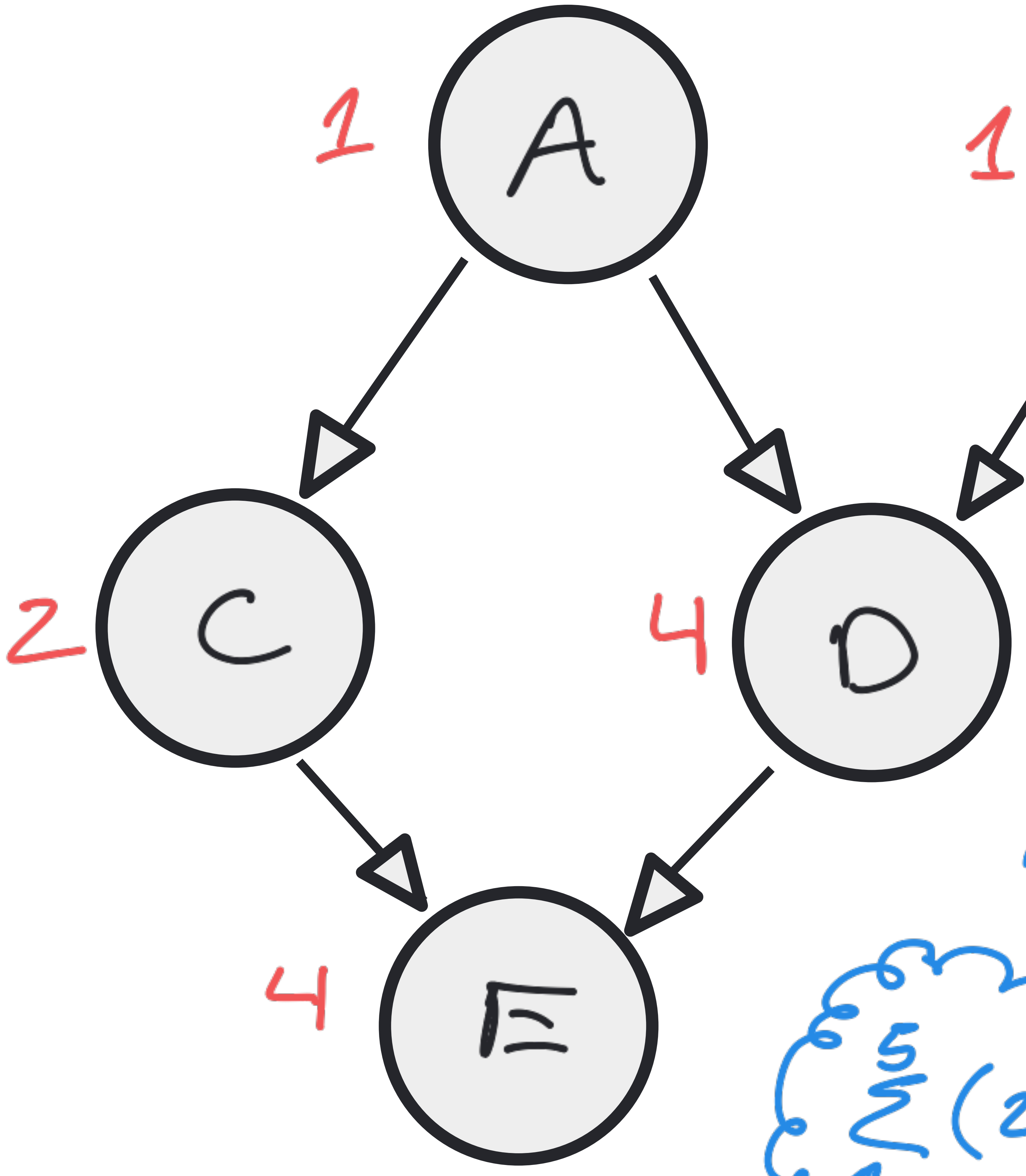
BAYES NETS: # of Parameters



1 param for A] - no parents
1 param for B]
2 params for C (1 parent)
4 params for D (2 parents)
4 params for E (2 parents)

Need 12 parameters for this Bayes Net under the Naive Bayes assumption

BAYES NETS: # of Parameters



All variables $\perp\!\!\!\perp$:
5 parameters

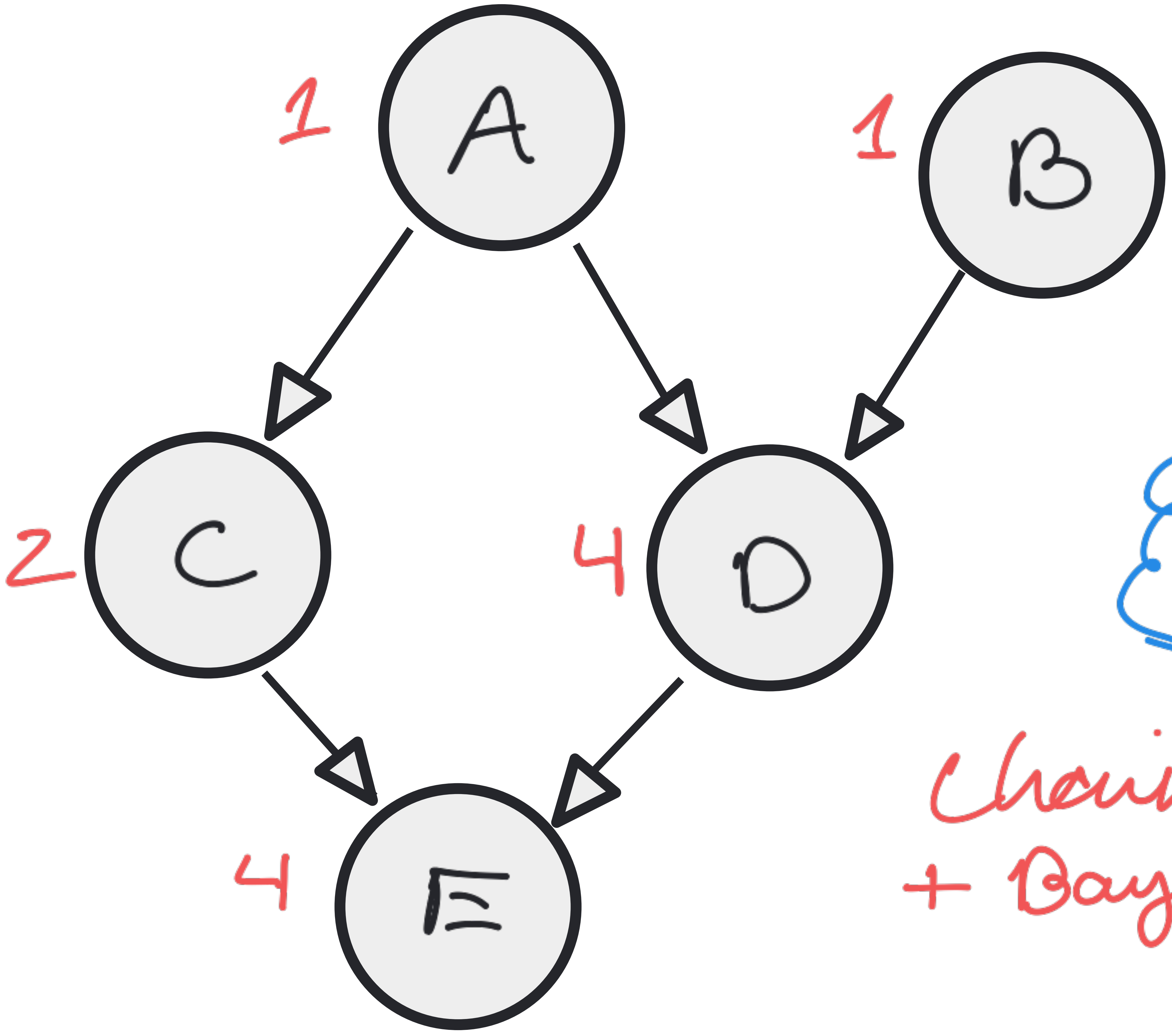
$$\sum_{1}^{5} (2-1)$$

Naive Bayes assumption:
12 parameters

No $\perp\!\!\!\perp$ assumption:
31 parameters!

$$\prod_{1}^{5} (2) - 1$$

BAYES NETS: # of Parameters



Applying the Bayes Net assumption:

Notation
$A \rightarrow A = \text{True}$
$\bar{A} \rightarrow A = \text{False}$

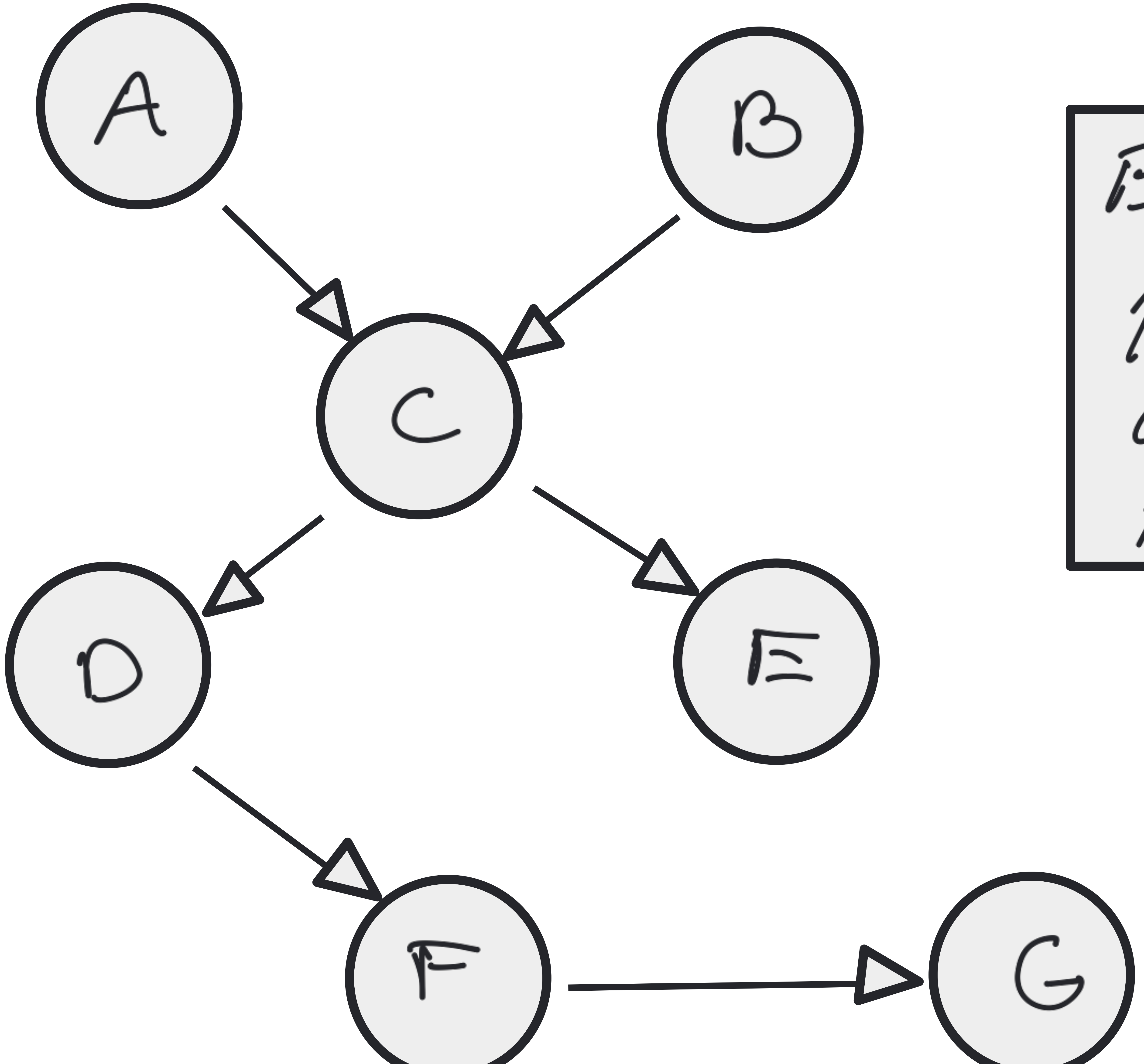
$$P(A, B, \bar{C}, D, \bar{E})$$

$$P(\bar{E} \mid D, \bar{C}, \cancel{B}, \cancel{A}) \cdot P(D \mid \cancel{\bar{C}}, B, A) \cdot P(\bar{C} \mid \cancel{B}, A) \cdot P(B \mid \cancel{A}) \cdot P(A)$$

Chain rule + Bayes net assumption

$$= P(\bar{E} \mid D, \bar{C}) P(D \mid B, A) P(\bar{C} \mid A) P(B) P(A)$$

BAYES NETS: # of Parameters



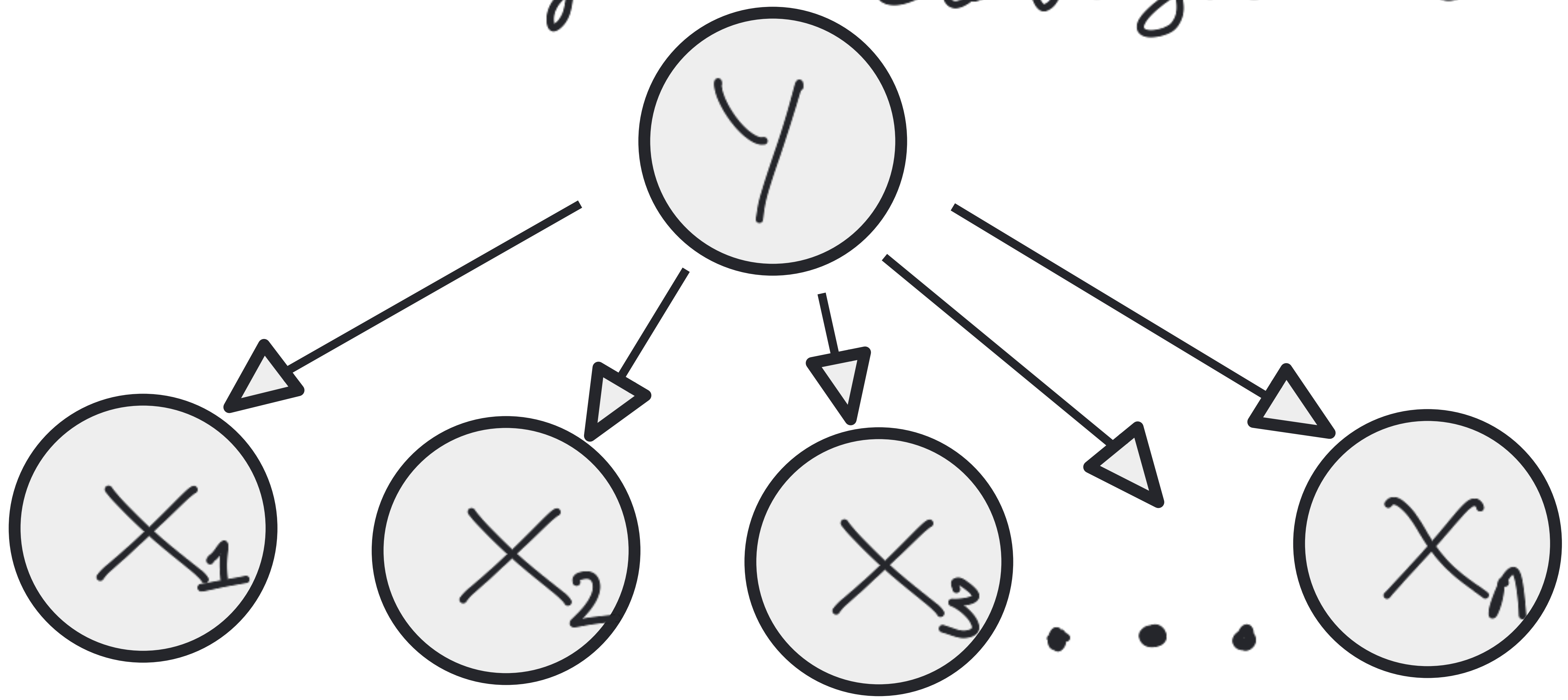
Exercise:

How many parameters are required in this Bayes Net?

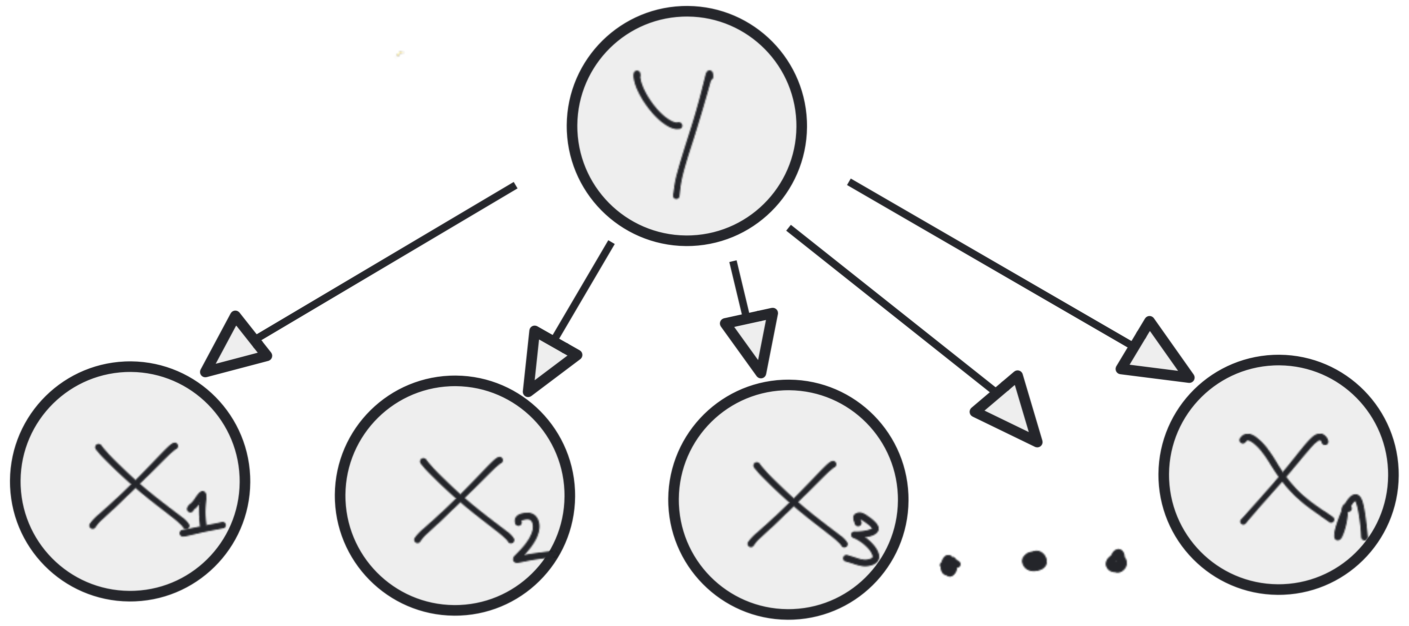
BAYES NETS : Naive Bayes Classification

★ <u>Naive Bayes assumption</u>:

features ($x_1, x_2, x_3, \ldots x_n$) are conditionally independent of each other given the final classification (Y)



BAYES NETS: Naive Bayes Classification



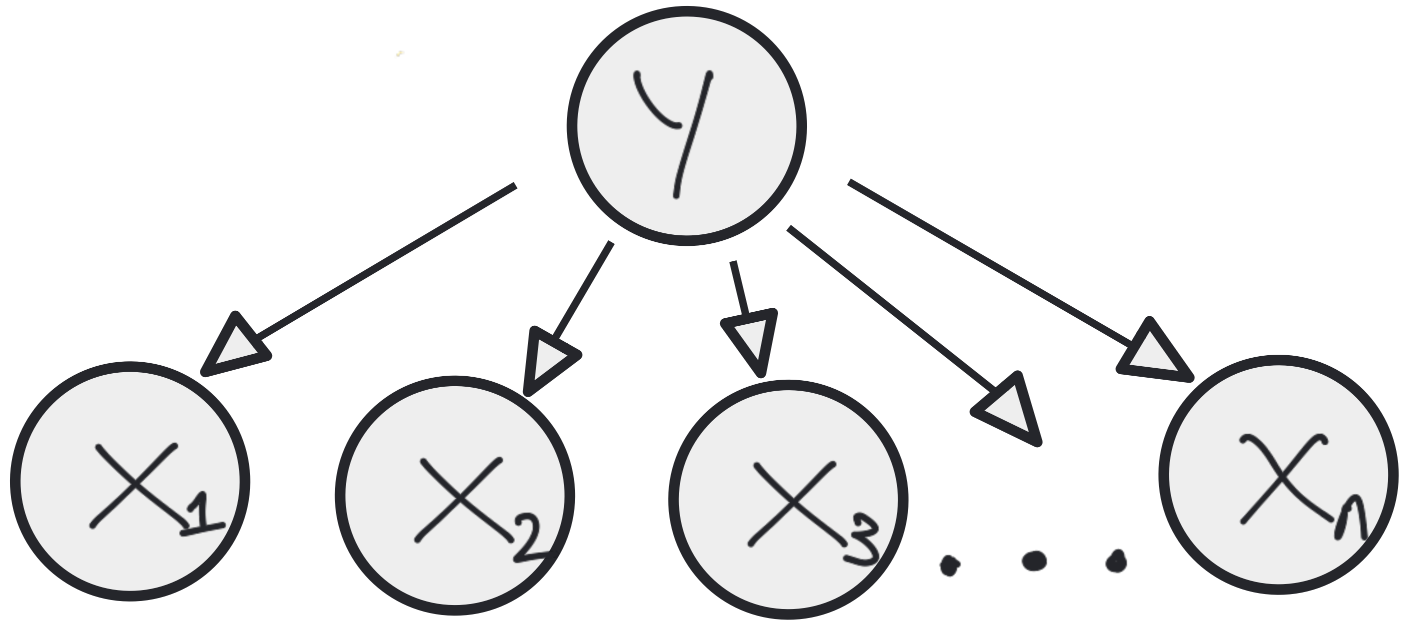
Example: Consider classifying people

person	likes toys	likes vegetables
child	✓	✗
adult	✗	✓
(Héctor)	(✓)	(✓)

1) If you know whether a point is an adult or a child, the likes toys or likes vegetables features are $\perp\!\!\!\perp$

2) If you don't know the classification, the features are related ($\not\perp\!\!\!\perp$)

BAYES NETS: Naive Bayes Classification



Example: Consider a point with known features but unknown classification:

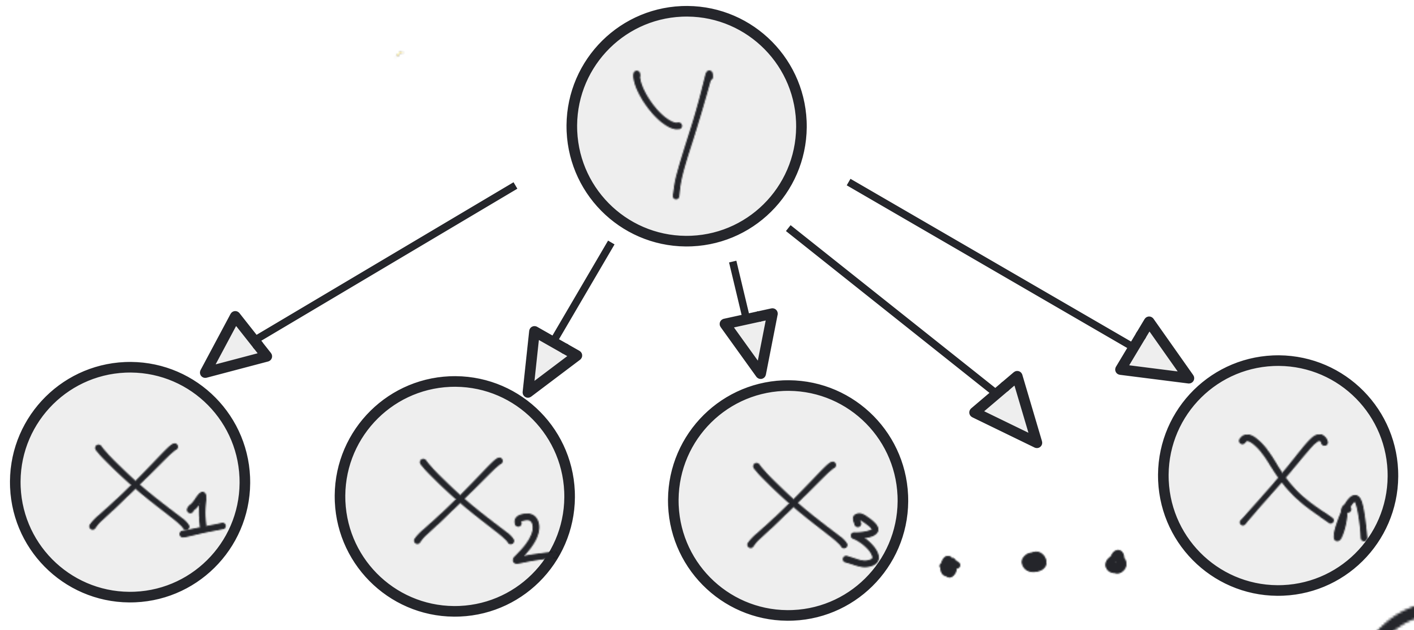
$$\{x_1=a, x_2=b, x_3=c\} = X$$

maximize

Goal: classify $X \longrightarrow \left[P(Y=y \mid \underbrace{X_1=a, X_2=b, X_3=c}_{\text{evidence}}) \right]$

posterior probability: probability after taking evidence into account

BAYES NETS: Naive Bayes Classification



Example: Consider a point with known features but unknown classification:

Goal: $\underset{y \in Y}{\operatorname{argmax}} \left[P(Y=y \mid X_1=a, X_2=b, X_3=c) \right]$

$$= \underset{y \in Y}{\operatorname{argmax}} \left[\frac{P(X_1=a, X_2=b, X_3=c \mid Y=y)\, P(Y=y)}{P(X=a, X=b, X=c)} \right]$$

the denominator does not change when y in Y=y changes, so it does not affect the max(·)

$$\sim \underset{y \in Y}{\operatorname{argmax}} \left[P(X_1=a, X_2=b, X_3=c \mid Y=y)\, P(Y=y) \right]$$

$$\Rightarrow \underset{y \in Y}{\operatorname{argmax}} \left[P(X_1=a \mid Y=y)\, P(X_2=b \mid Y=y)\, P(X_3=c \mid Y=y)\, P(Y=y) \right]$$

prior probability must be known beforehand

BAYES NETS: Model Selection

1) Start with several possible models (m_i) that could explain the data or evidence

2) Pick the model (m_i) that maximizes

$$P(m_i \mid data) = \frac{P(data \mid m_i)\,P(m_i)}{P(data)}$$

$$\propto P(data \mid m_i)\,P(m_i)$$

(the denominator does not change across m_i)

BAYES NETS: Model Selection

2010 Final

2 coins:

- 1 fair – $P(H) = 0.5$
- 1 biased – $P(H) = 1$

2 models: (equal probability)

- M_1: draw a coin, toss it once, put it back, repeat twice
- M_2: draw a coin, toss it three times

Data: HHT

want: $\underset{i \in \{1,2\}}{\operatorname{argmax}} \left\{ P(m_i \mid HHT) \right\}$

BAYES NETS: Model Selection

<u>2010 Final</u>

$$P(M_1 \mid HHT) = \frac{P(HHT \mid M_1)\,P(M_1)}{P(HHT)}$$

$$= \frac{P(H \mid M_1)\,P(H \mid M_1)\,P(T \mid M_1)\,P(M_1)}{\cancel{P(HHT)}}$$ not needed

$$P(H \mid M_1) = P(H \mid \text{fair})\,P(\text{fair}) + P(H \mid \text{biased})\,P(\text{biased})$$

$$= 0.75$$

$$P(T \mid M_1) = \tfrac{1}{2} P(T \mid \text{fair}) + \cancelto{0}{\tfrac{1}{2} P(T \mid \text{biased})}$$

$$= 0.25$$

$$(\text{also: } 1 - P(H \mid M_1))$$

$$\propto (0.75)(0.75)(0.25)(0.5)$$

$$= \left(\tfrac{3}{4}\right)^2 \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right)$$

BAYES NETS: Model Selection

<u>2010 Final</u>

$$P(M_2 \mid HHT) = \frac{P(HHT \mid M_2)\,P(M_2)}{P(HHT)}$$

$$= \frac{\left[P(HHT \mid \text{fair})P(\text{fair}) + P(HHT \mid \text{biased})P(\text{biased})\right]P(M_2)}{\cancel{P(HHT)}}$$ not needed

$$\propto \left[\left(\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)\right)\left(\tfrac{1}{2}\right) + (0)\left(\tfrac{1}{2}\right)\right]\left(\tfrac{1}{2}\right)$$

$$= \left(\tfrac{1}{2}\right)^5$$

BAYES NETS: Model Selection

2010 Final

$$P(M_1 \mid HHT) = \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)\left(\frac{1}{2}\right)$$

$$P(M_2 \mid HHT) = \left(\frac{1}{2}\right)^5$$

$$P(M_1 \mid HHT) > P(M_2 \mid HHT)$$

☆ M_1 explains the data better

BAYES NETS: Gold Star Ideas

- ☆ Bayes Net assumption
  (variable $\perp\!\!\!\perp$ nondescendants | parents)

- ☆ Naive Bayes assumption
  (features $\perp\!\!\!\perp$ each other | classification)