Factor Graphs

Discussion:

Practice with Factor Graphs and Belief Propagation

ECE/CS 498 DS

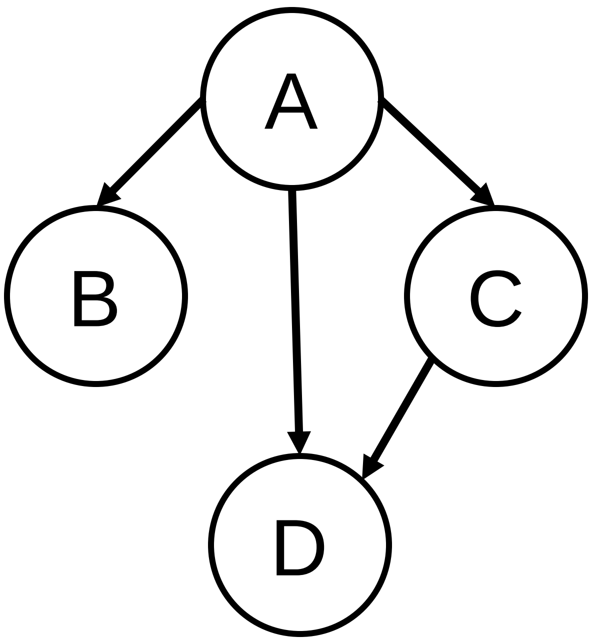
University of Illinois

Agenda

- Conversion of Bayesian network to factor graph (FG)
- Calculating joint probability in a FG
- Calculating marginal probability with belief propagation
- Calculating conditional probability with belief propagation

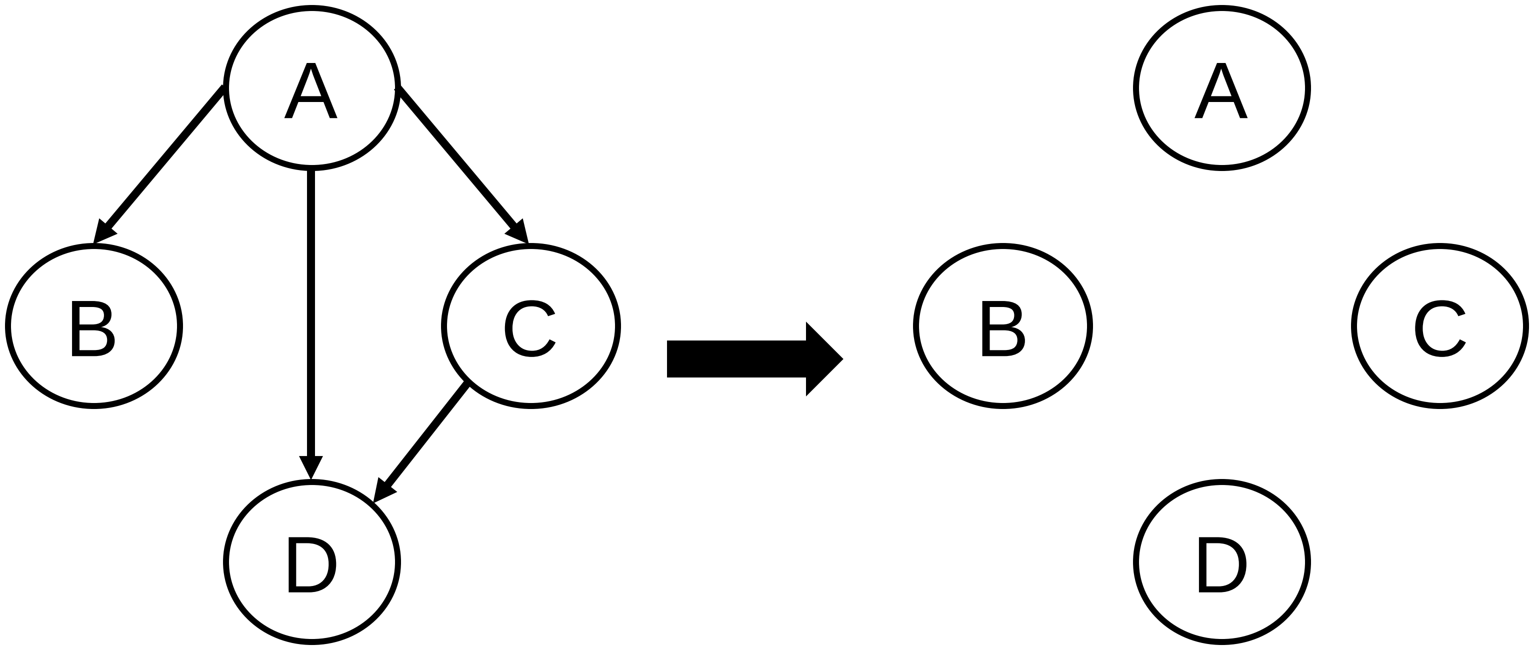
Converting Bayesian Network to Factor Graph

Bayesian Network Example



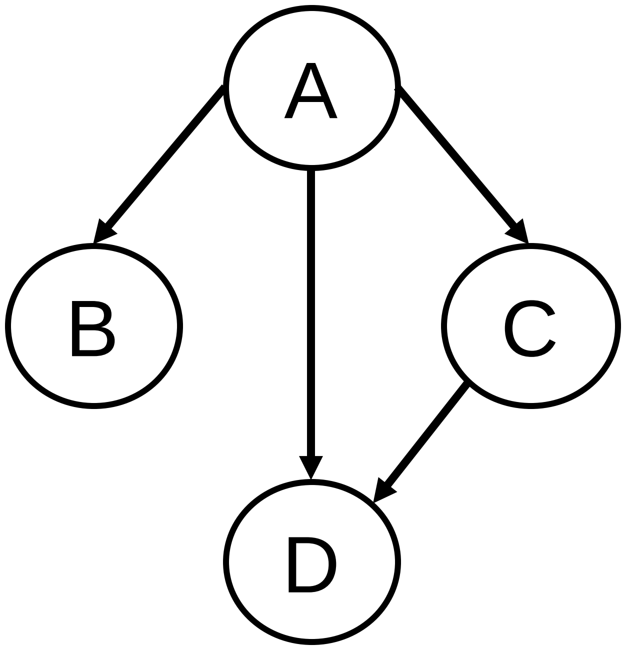
- Suppose that we have a Bayesian network as shown to the left
- Every variable is binary, and takes on a value from $\{0, 1\}$
- **How can we convert this to a factor graph?**

Conversion to FG: Draw Variables



- First, we need to identify all the variables that will be drawn in the FG

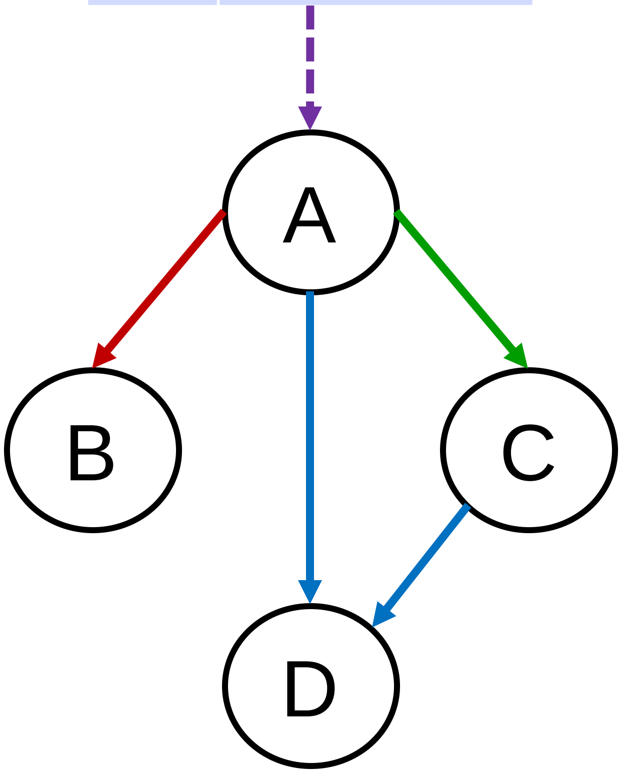
Conversion to FG: Identify Factor Functions



- Next, we need to identify what the factor functions will be
- To do this, think about the CPTs of the Bayesian network - how do we encode the **variable dependencies**?
  - CPTs describe dependencies of a node on its parents

Conversion to FG: CPTs

A	$P(A)$
0	0.9



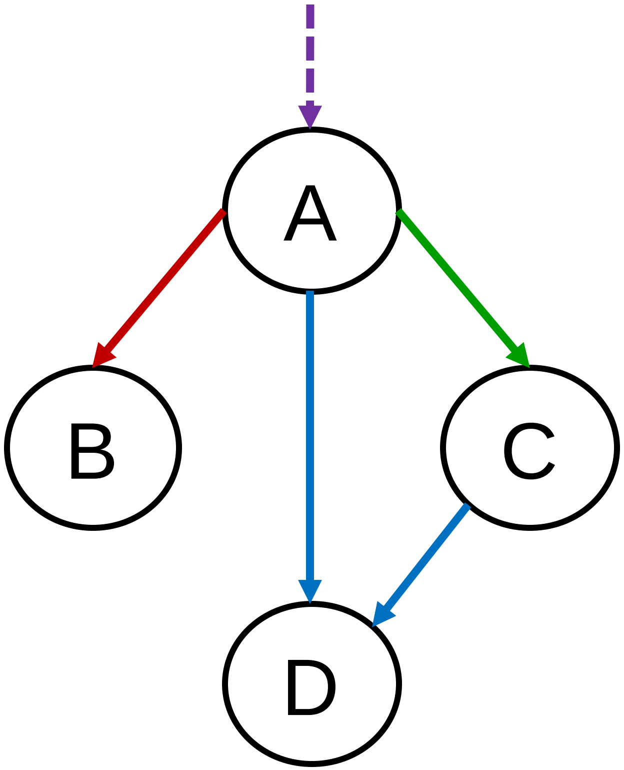
$CPTs$

A	$P(B=0\|A)$
0	0.6
1	0.3

A	$P(C=0\|A)$
0	0.8
1	0.5

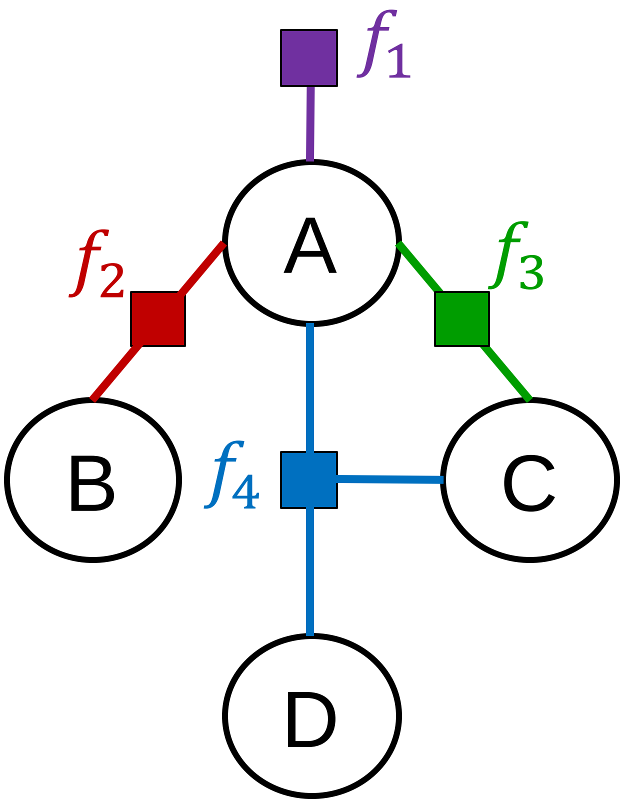
A	C	$P(D=0\|A,C)$
0	0	0.1
0	1	0.4
1	0	0.2
1	1	0.7

Conversion to FG: CPTs



- Using the CPTs, we can create factor functions

Conversion to FG: Factor Functions



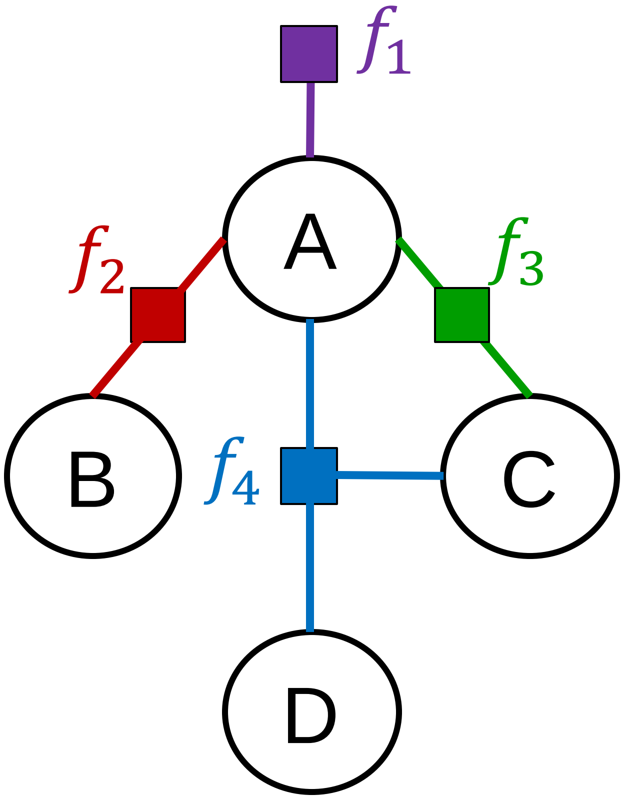
f_1 represents a prior probability ($P(A)$)

A	$P(A)$
0	0.9

A	$f_1(A)$
0	9
1	1

- Note that while CPTs only list the required/independent parameters, factor functions should consider all possible parameters (even dependent ones)
- We multiply by probabilities by 10 for convenience in calculations

Conversion to FG: Factor Functions



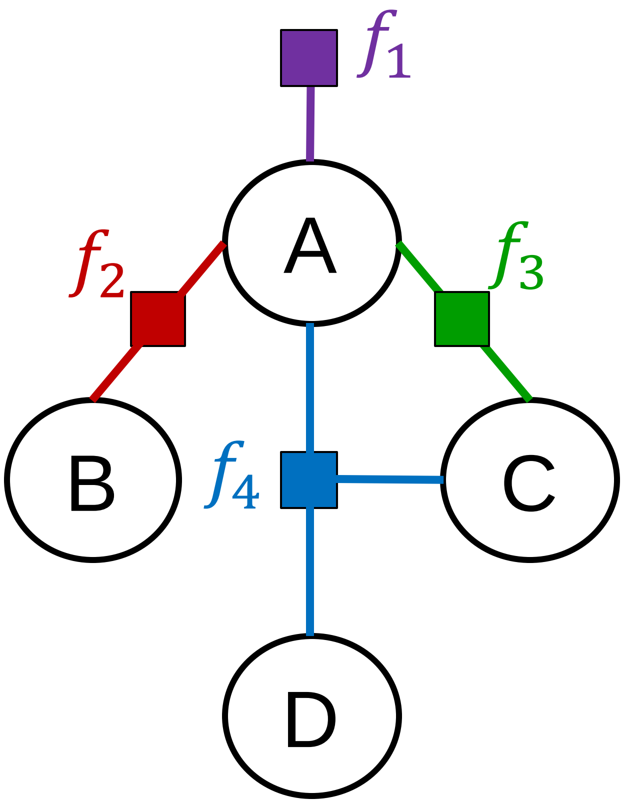
f_2 represents a conditional probability ($P(B|A)$)

A	$P(B=0\|A)$
0	0.6
1	0.3

⬇

A	B	$f_2(A,B)$
0	0	6
0	1	4
1	0	3
1	1	7

Conversion to FG: Factor Functions



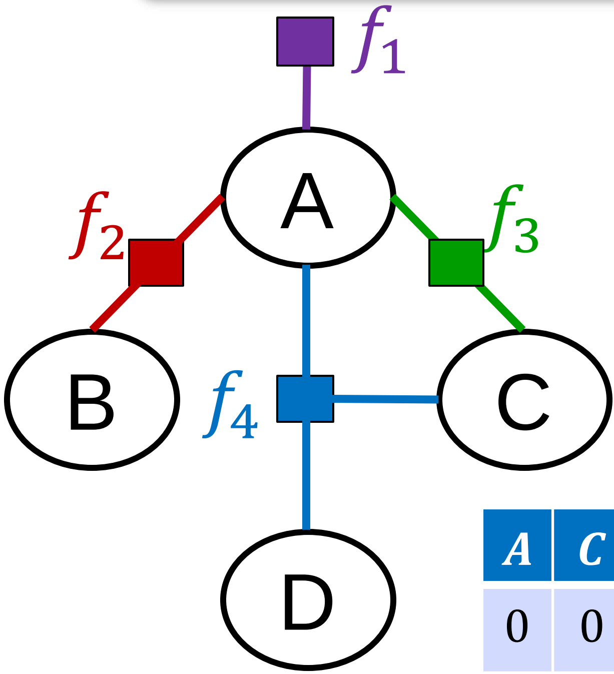
f_3 represents a conditional probability ($P(C|A)$)

A	$P(C = 0|A)$
0	0.8
1	0.5

⬇

A	C	$f_3(A, C)$
0	0	8
0	1	2
1	0	5
1	1	5

Conversion to FG: Factor Functions

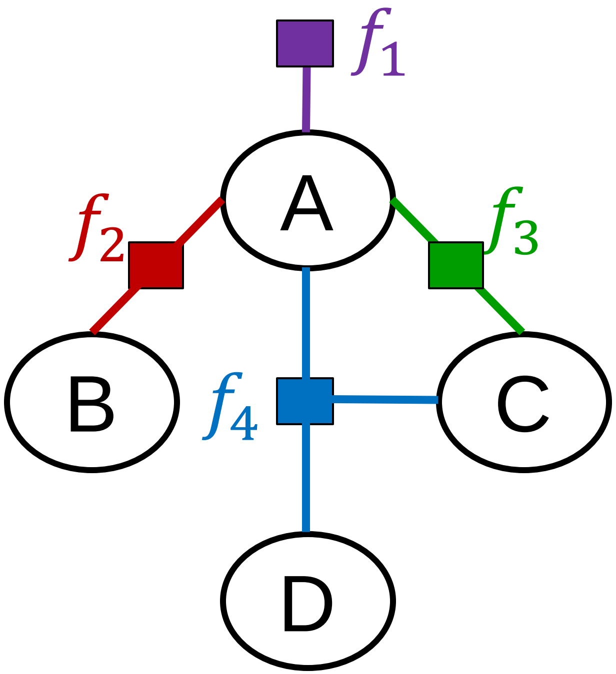


f_4 represents a conditional probability ($P(D|A,C)$)

A	C	$P(D=0\|A,C)$
0	0	0.1
0	1	0.4
1	0	0.2
1	1	0.7

A	C	D	$f_4(A,C,D)$
0	0	0	1
0	0	1	9
0	1	0	4
0	1	1	6
1	0	0	2
1	0	1	8
1	1	0	7
1	1	1	3

Final Factor Graph



A	$f_1(A)$
0	9
1	1

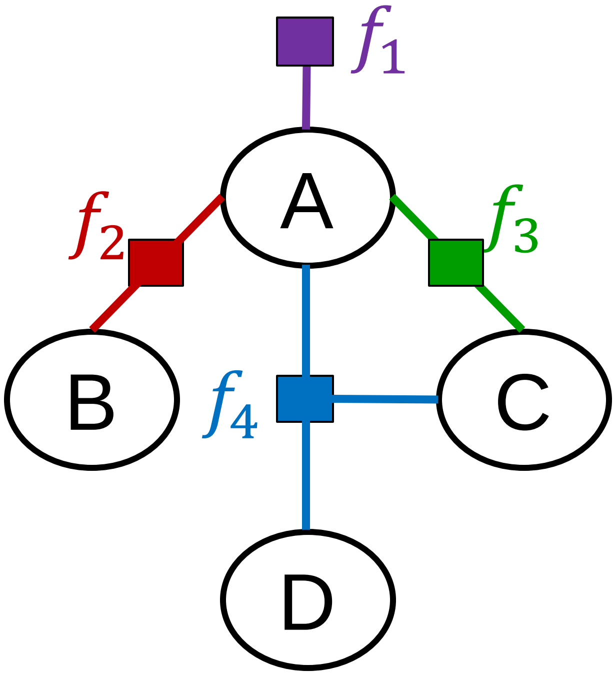
A	B	$f_2(A,B)$
0	0	6
0	1	4
1	0	3
1	1	7

A	C	$f_3(A,C)$
0	0	8
0	1	2
1	0	5
1	1	5

A	C	D	$f_4(A,C,D)$
0	0	0	1
0	0	1	9
0	1	0	4
0	1	1	6
1	0	0	2
1	0	1	8
1	1	0	7
1	1	1	3

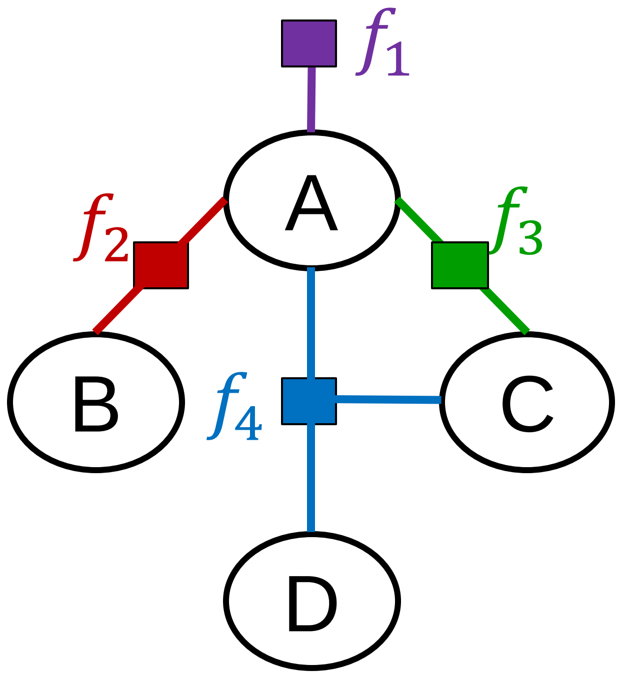
Calculating Joint Probability in a Factor Graph

Calculating Joint Probability



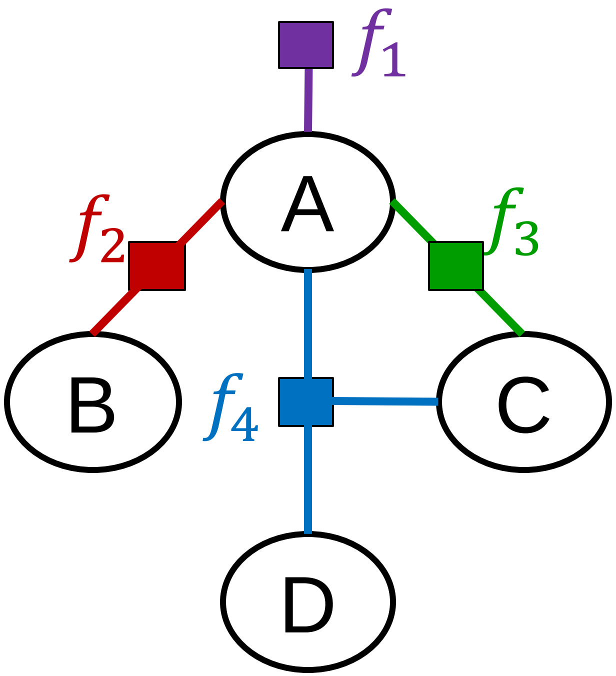
- **Given the factor graph to the left, how do we calculate the joint probability $\boldsymbol{P(A=1, B=0, C=1, D=1)}$?**
- We need to multiply the factor functions together, and then normalize the result

Multiplying Factor Functions



$$P(A=1,B=0,C=1,D=1)$$
$$\propto f_1(A=1)f_2(A=1,B=0)f_3(A=1,C=1)f_4(A=1,C=1,D=1)$$
$$=(1)*(3)*(5)*(3)$$
$$=45$$

Calculating Partition Function



- The partition function Z is simply the sum of all possible joint factor products
- Since the factor functions represent affinities (and not necessarily probabilities), we need to normalize their product to calculate the joint probability
- Refer to “additional slides” section for how we calculated Z below

$$Z = \sum_{a,b,c,d \in \{0,1\}} f_1(A=a) f_2(A=a, B=b) f_3(A=a, C=c) f_4(A=a, C=c, D=d)$$

$$= 10{,}000$$

Reporting Joint Probability

- Thus,

$$\boldsymbol{P(A=1, B=0, C=1, D=1)}$$

$$=\frac{45}{\mathrm{Z}}$$

$$=\frac{45}{10000}$$

$$=0.0045$$

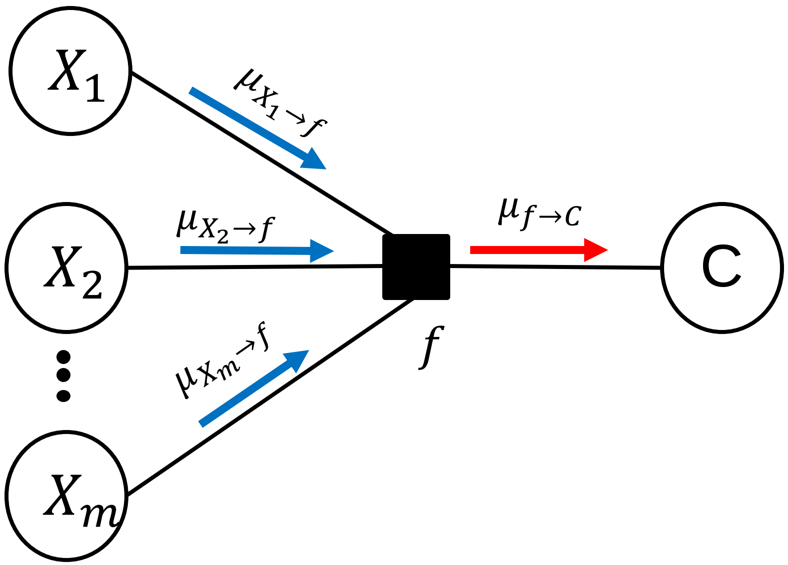
Calculating Marginal Probability using Belief Propagation

Motivation: Why Belief Propagation?

- Belief propagation can greatly reduce the number of required computations for calculating marginal and conditional probabilities with a factor graph
- Calculating all the joint probabilities grows exponentially with the number of variables
  - Refer to the tables in the “Additional Slides” section for an example of this!
- By passing messages across variables and factors, we can save time by not performing any redundant calculations
- We will demonstrate using a slightly modified version of our earlier factor graph (without loops…)

Belief Propagation – Message Passing

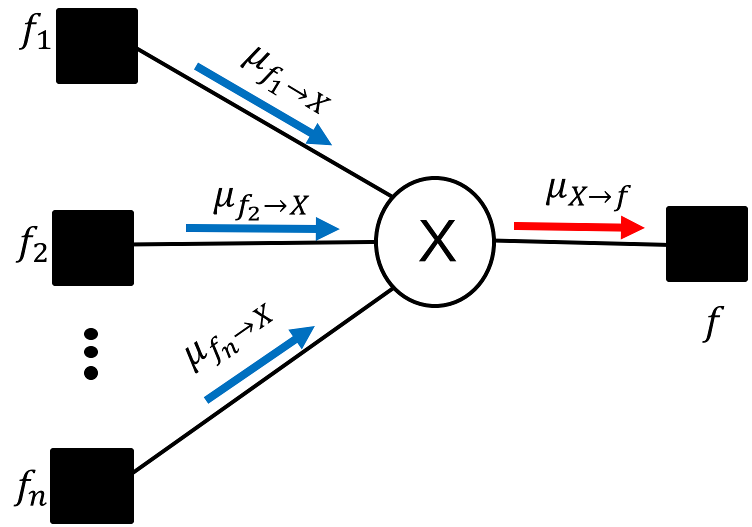
Message from Factor to Variable



$$\mu_{f \to C}(C) = \sum_{X_1, X_2, \ldots, X_m} f(C, X_1, \ldots, X_m) \prod_{i=1}^{m} \mu_{X_i \to f}(X_i)$$

Message from factor to variable: Product of all incoming messages and factor, sum out previous variables

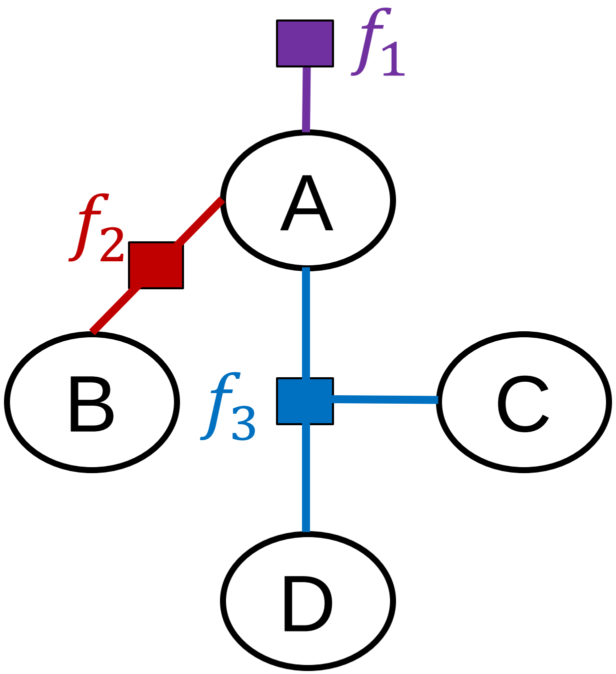
Message from Variable to Factor



$$\mu_{X \to f}(X) = \prod_{i=1}^{n} \mu_{f_i \to X}(X)$$

Message from variable to factor: Product of all incoming messages

New Factor Graph

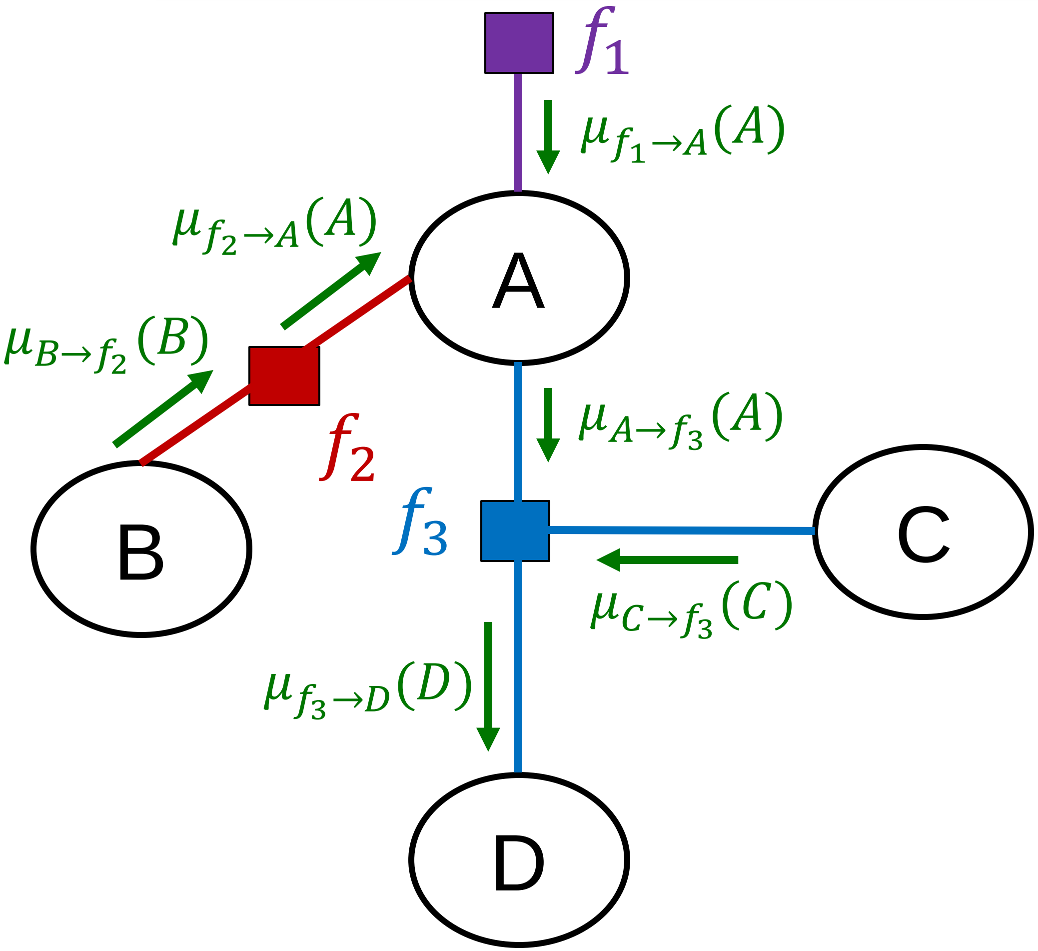


A	$f_1(A)$
0	9
1	1

A	B	$f_2(A,B)$
0	0	6
0	1	4
1	0	3
1	1	7

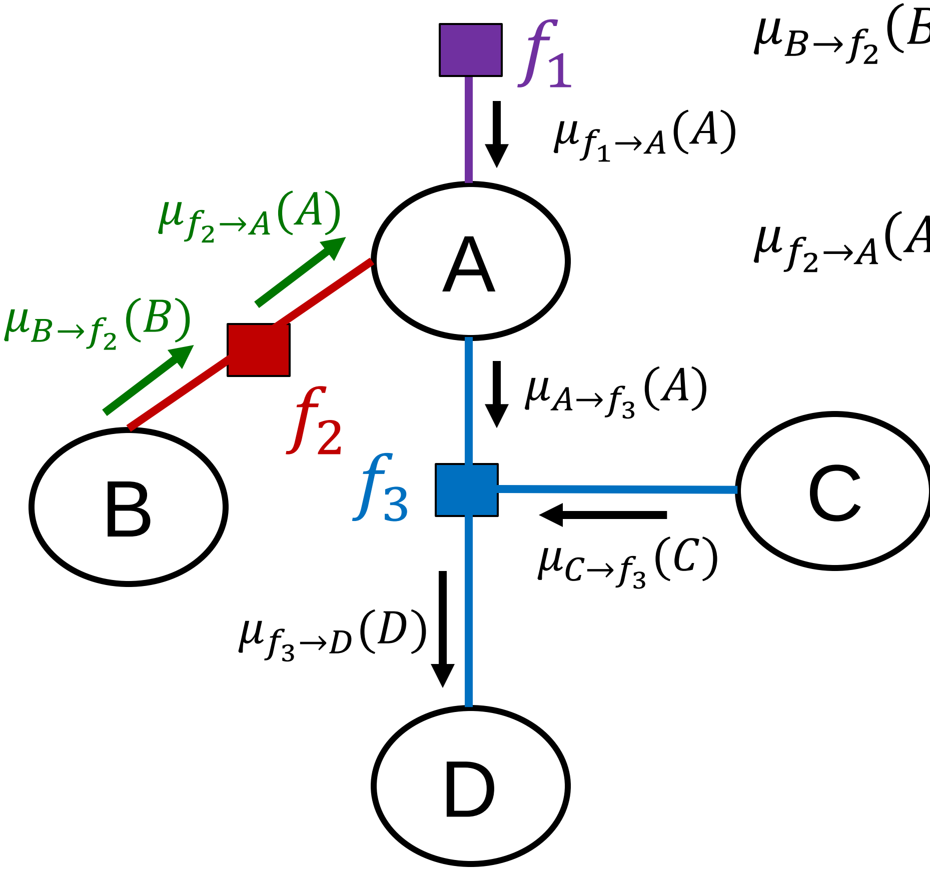
A	C	D	$f_3(A,C,D)$
0	0	0	1
0	0	1	9
0	1	0	4
0	1	1	6
1	0	0	2
1	0	1	8
1	1	0	7
1	1	1	3

Calculating Marginal Probability



- **Given the factor graph to the left, how do we calculate the marginal probability $\boldsymbol{P(D=0)}$?**
- The first step is to identify all messages that are being sent to D.

Calculating Marginal Probability



$$\mu_{B\to f_2}(B) = \begin{bmatrix}1\\1\end{bmatrix} \begin{matrix}(B=0)\\(B=1)\end{matrix}$$

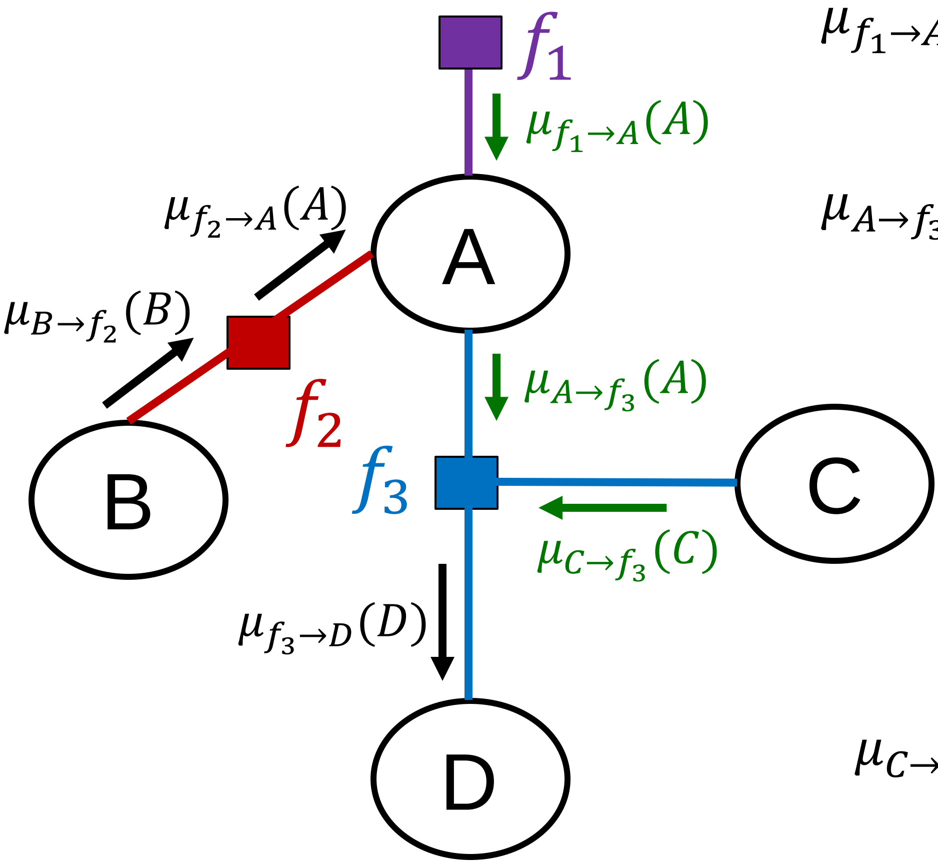
$$\mu_{f_2\to A}(A) = \sum_{B\in\{0,1\}} f_2(A,B)\times\mu_{B\to f_2}(B)$$

$$= f_2(A,B=0)\times\mu_{B\to f_2}(B=0) + f_2(A,B=1)\times\mu_{B\to f_2}(B=1)$$

$$= \begin{bmatrix}6\\3\end{bmatrix} \begin{matrix}(A=0)\\(A=1)\end{matrix} \times 1 + \begin{bmatrix}4\\7\end{bmatrix} \begin{matrix}(A=0)\\(A=1)\end{matrix} \times 1$$

$$= \begin{bmatrix}10\\10\end{bmatrix} \begin{matrix}(A=0)\\(A=1)\end{matrix}$$

Calculating Marginal Probability



$$\mu_{f_1\to A}(A) = \begin{bmatrix} 9 \\ 1 \end{bmatrix} \begin{matrix} (A=0) \\ (A=1) \end{matrix}$$

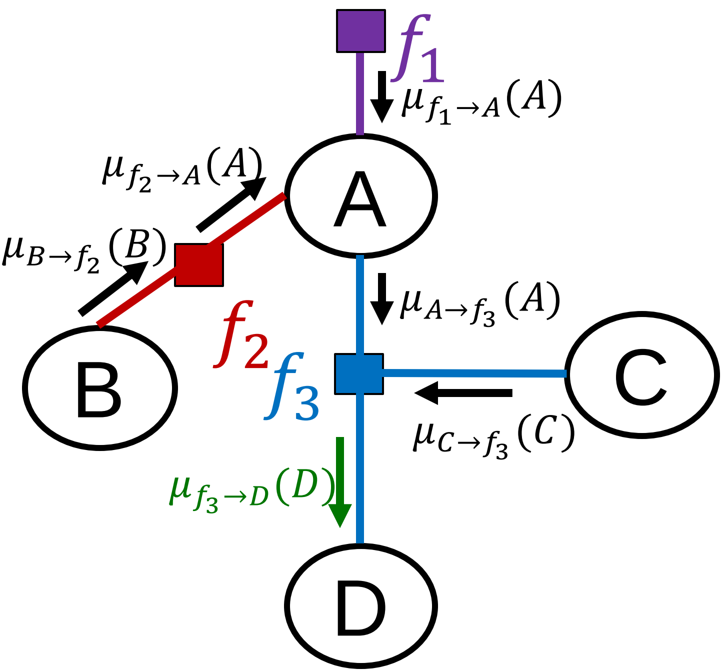
$$\mu_{A\to f_3}(A) = \mu_{f_2\to A}(A) \times \mu_{f_1\to A}(A)$$

$$= \begin{bmatrix} 10 \\ 10 \end{bmatrix} \begin{matrix} (A=0) \\ (A=1) \end{matrix} \odot \begin{bmatrix} 9 \\ 1 \end{bmatrix} \begin{matrix} (A=0) \\ (A=1) \end{matrix}$$

$$= \begin{bmatrix} 90 \\ 10 \end{bmatrix} \begin{matrix} (A=0) \\ (A=1) \end{matrix}$$

$$\mu_{C\to f_3}(C) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{matrix} (C=0) \\ (C=1) \end{matrix}$$

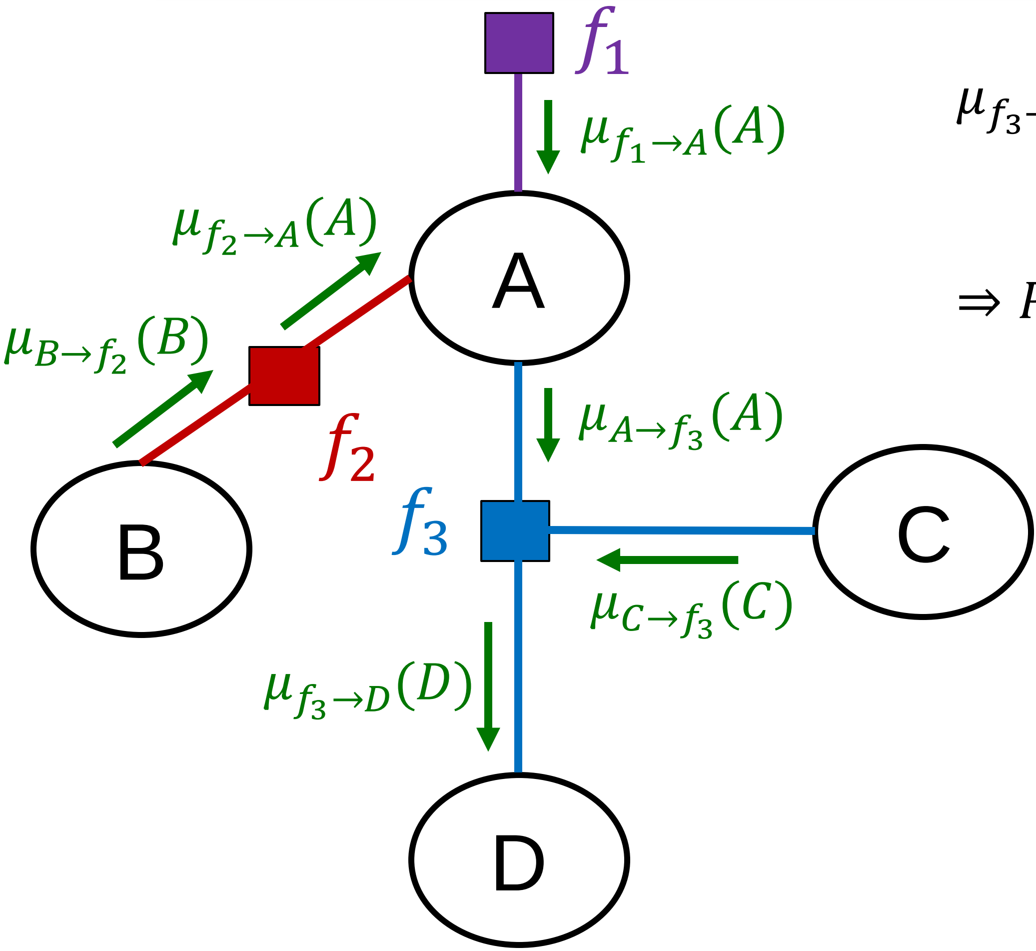
Calculating Marginal Probability



$$\mu_{f_3\to D}(D)$$
$$= \sum_{A\in\{0,1\},C\in\{0,1\}} f_3(A,C,D)\times\mu_{A\to f_3}(A)\times\mu_{C\to f_3}(C)$$

$$\begin{aligned}
&= f_3\,(A=0,C=0,D)\times\mu_{A\to f_3}(A=0)\times\mu_{C\to f_3}(C=0)\\
&+f_3(A=0,C=1,D)\times\mu_{A\to f_3}(A=0)\times\mu_{C\to f_3}(C=1)\\
&+f_3(A=1,C=0,D)\times\mu_{A\to f_3}(A=1)\times\mu_{C\to f_3}(C=0)\\
&+f_3(A=1,C=1,D)\times\mu_{A\to f_3}(A=1)\times\mu_{C\to f_3}(C=1)\\
&= \begin{bmatrix}1\\9\end{bmatrix}\begin{matrix}(D=0)\\(D=1)\end{matrix}\times 90\times 1+\begin{bmatrix}4\\6\end{bmatrix}\begin{matrix}(D=0)\\(D=1)\end{matrix}\times 90\times 1\\
&+\begin{bmatrix}2\\8\end{bmatrix}\begin{matrix}(D=0)\\(D=1)\end{matrix}\times 10\times 1+\begin{bmatrix}7\\3\end{bmatrix}\begin{matrix}(D=0)\\(D=1)\end{matrix}\times 10\times 1\\
&= \begin{bmatrix}540\\1460\end{bmatrix}\begin{matrix}(D=0)\\(D=1)\end{matrix}
\end{aligned}$$

Calculating Marginal Probability



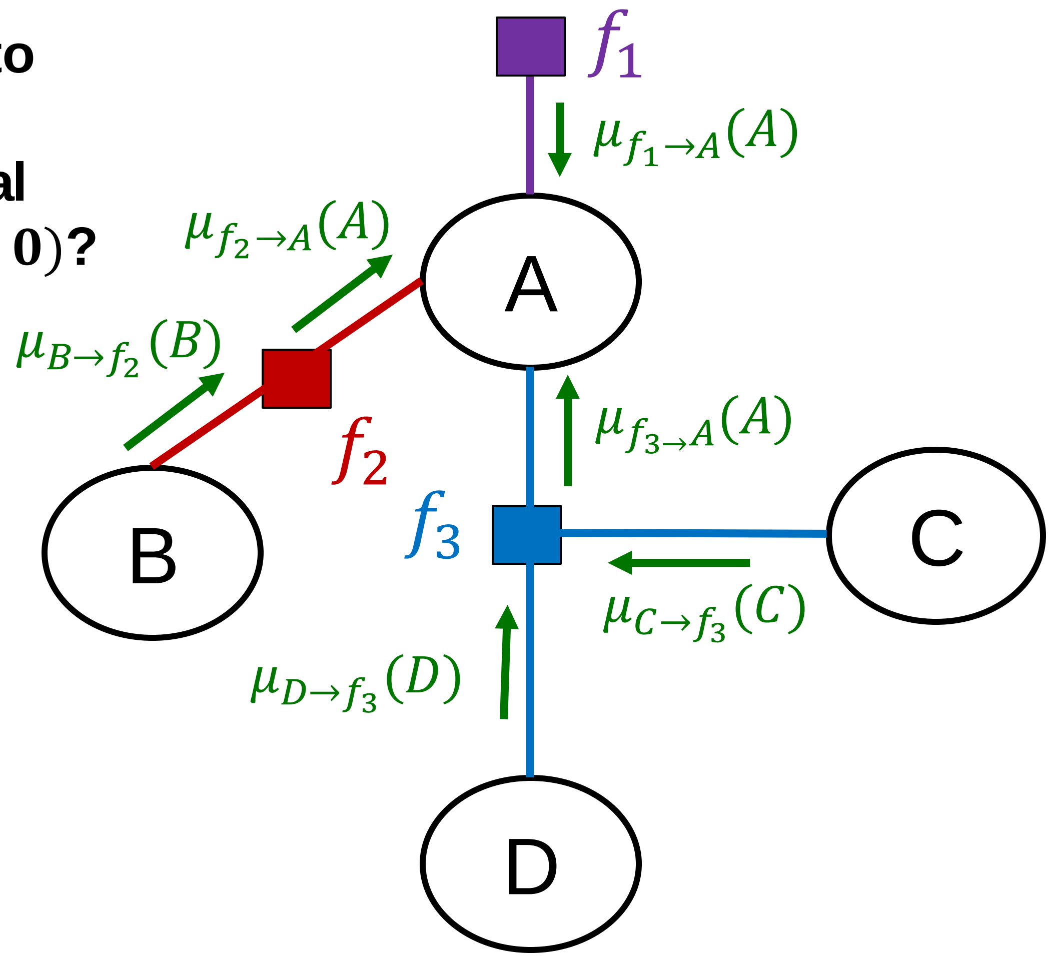
$$\mu_{f_3 \to D}(D) = \begin{bmatrix} 540 \\ 1460 \end{bmatrix} \begin{matrix} (D = 0) \\ (D = 1) \end{matrix}$$

$$\Rightarrow P(D = 0) = \frac{540}{540+1460} = \frac{540}{2000} = 0.27$$

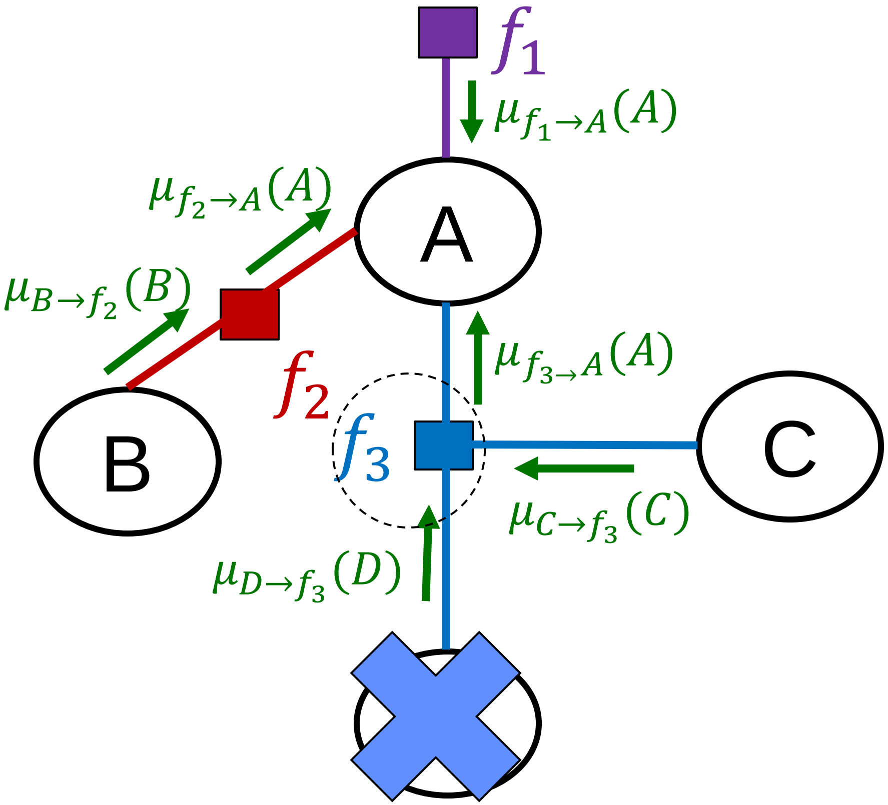
Calculating Conditional Probability using Belief Propagation

Calculating Conditional Probability

- **Given the factor graph to the right, how do we calculate the conditional probability $P(A=0|D=0)$?**



Calculating Conditional Probability

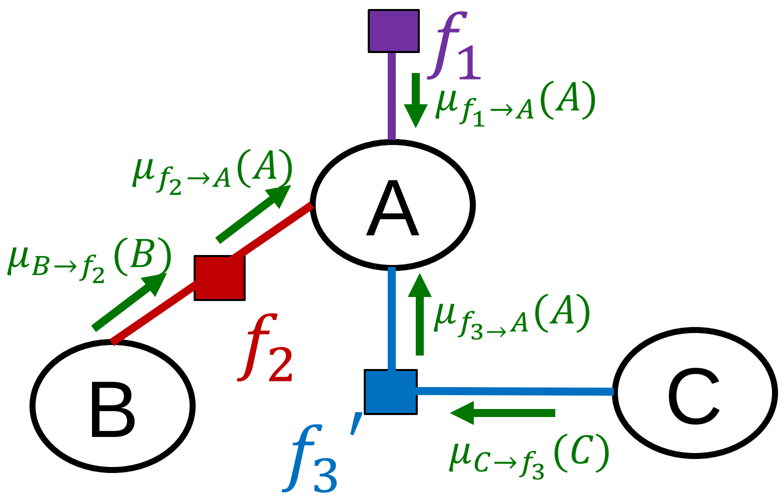


On observing Node D, we are going to make corresponding changes in the factor graph:

- Remove node D from factor graph
- Modify the factor function connecting to node D (i.e. factor function f_3)

Calculating Conditional Probability

On observing Node $D = 0$,
We reserve the entries with $D = 0$
in factor function f_3

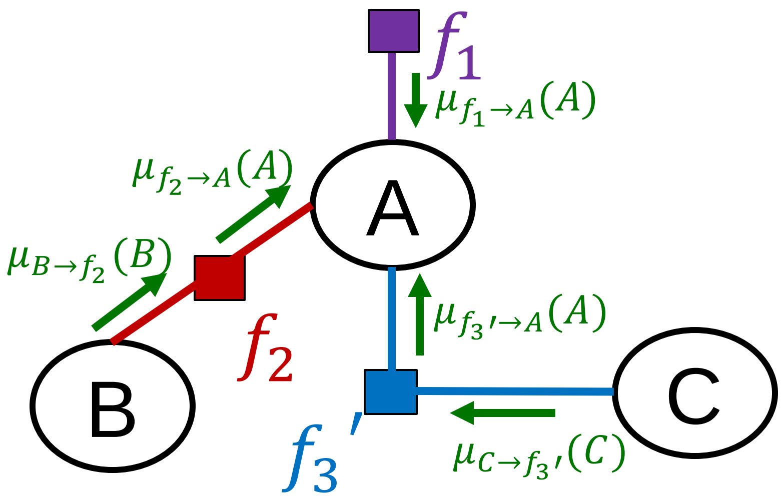


A	C	D	$f_3(A, C, D)$
0	0	0	1
~~0~~	~~0~~	~~1~~	~~9~~
0	1	0	4
~~0~~	~~1~~	~~1~~	~~6~~
1	0	0	2
~~1~~	~~0~~	~~1~~	~~8~~
1	1	0	7
~~1~~	~~1~~	~~1~~	~~3~~

Calculating Conditional Probability

On observing Node $D = 0$,
We reserve the entries with $D = 0$ in factor function f_3 and obtain new $f_3{}'$ to the right

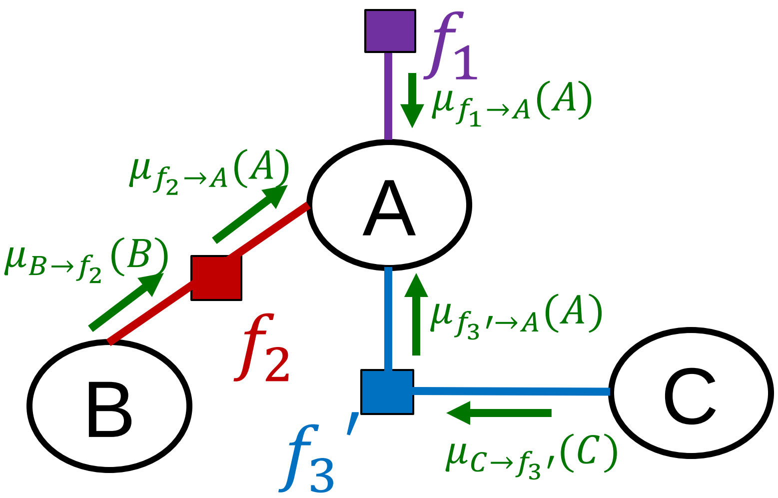
A	C	(D)	$f_3{}'(A,C)$
0	0	0	1
0	1	0	4
1	0	0	2
1	1	0	7



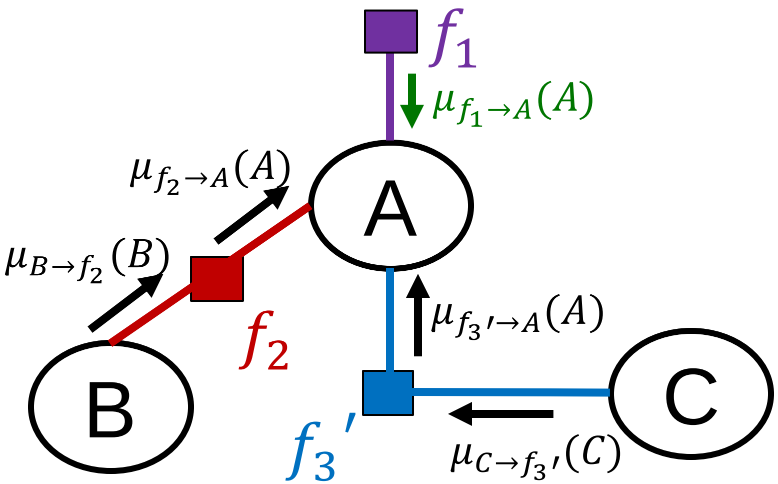
Calculating Conditional Probability

On observing Node $D = 0$,
We reserve the entries with $D = 0$ in factor function f_3 and obtain new $f_3{}'$ to the right

A	C	$f_3{}'(A, C)$
0	0	1
0	1	4
1	0	2
1	1	7



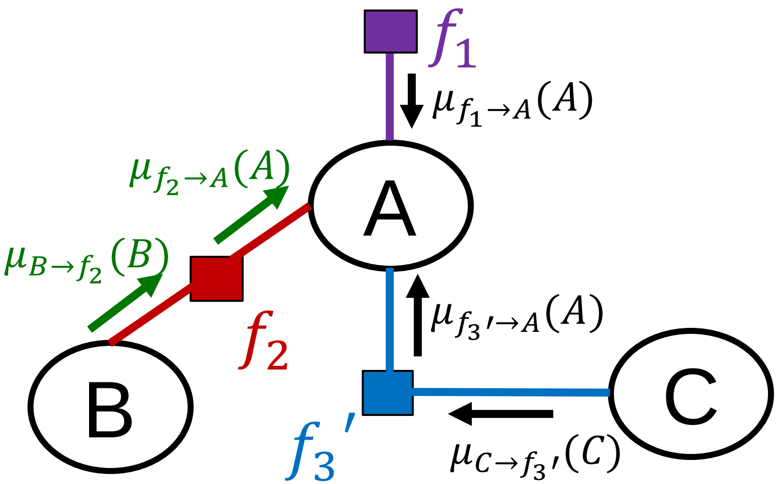
Calculating Conditional Probability



$$\mu_{f_1 \to A}(A) = \begin{bmatrix} 9 \\ 1 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix}$$

A	$f_1(A)$
0	9
1	1

Calculating Conditional Probability



$$\mu_{B \to f_2}(B) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{matrix} (B = 0) \\ (B = 1) \end{matrix}$$

$$\mu_{f_2 \to A}(A) = \sum_{B \in \{0,1\}} f_2(A, B) \times \mu_{B \to f_2}(B)$$

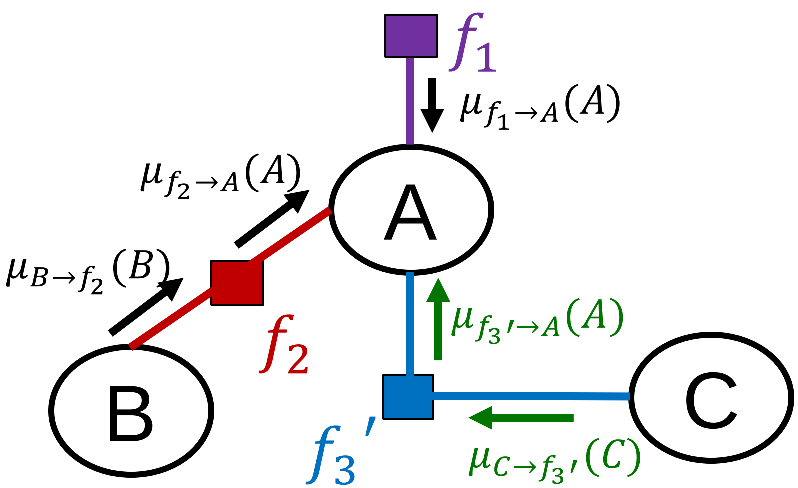
$$= f_2(A, B = 0) \times \mu_{B \to f_2}(B = 0) + f_2(A, B = 1) \times \mu_{B \to f_2}(B = 1)$$

$$= \begin{bmatrix} 6 \\ 3 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix} \times 1 + \begin{bmatrix} 4 \\ 7 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix} \times 1$$

$$= \begin{bmatrix} 10 \\ 10 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix}$$

A	B	$f_2(A, B)$
0	0	6
0	1	4
1	0	3
1	1	7

Calculating Conditional Probability



A	C	$f_3{}'(A, C)$
0	0	1
0	1	4
1	0	2
1	1	7

$$\mu_{C \to f_3{}'}(C) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{matrix} (C = 0) \\ (C = 1) \end{matrix}$$

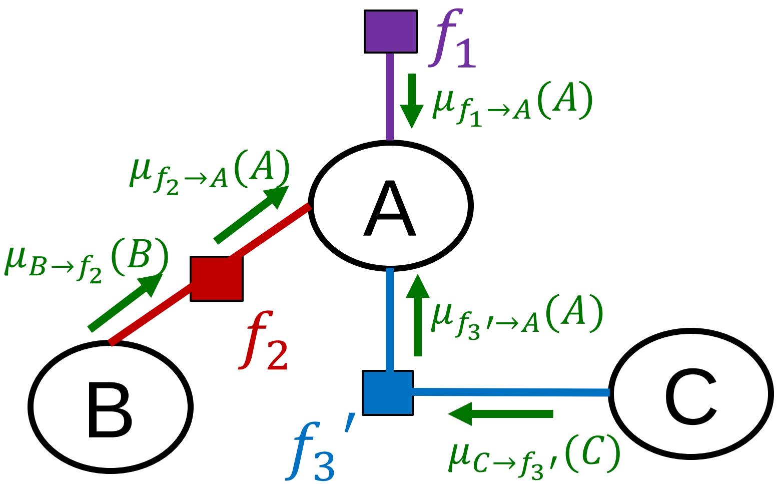
$$\mu_{f_3{}' \to A}(A) = \sum_{C \in \{0,1\}} f_3{}'(A, C) \times \mu_{C \to f_3{}'}(C)$$

$$= f_3{}'(A, C = 0) \times \mu_{C \to f_3{}'}(C = 0)$$
$$+ f_3{}'(A, C = 1) \times \mu_{C \to f_3{}'}(C = 1)$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix} \times 1 + \begin{bmatrix} 4 \\ 7 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix} \times 1$$

$$= \begin{bmatrix} 5 \\ 9 \end{bmatrix} \begin{matrix} (A = 0) \\ (A = 1) \end{matrix}$$

Calculating Conditional Probability



$$P(A|D=0) = \frac{1}{Z}(\mu_{f_1 \to A} \times \mu_{f_2 \to A} \times \mu_{f_3' \to A})$$

$$= \frac{1}{Z}\left(\begin{bmatrix} 9 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 10 \\ 10 \end{bmatrix} \odot \begin{bmatrix} 5 \\ 9 \end{bmatrix}\right)$$

$$= \frac{1}{Z}\begin{bmatrix} 450 \\ 90 \end{bmatrix}$$

$$= \frac{1}{540}\begin{bmatrix} 450 \\ 90 \end{bmatrix}$$

$$= \begin{bmatrix} 5/6 \\ 1/6 \end{bmatrix} \begin{matrix} (A=0) \\ (A=1) \end{matrix}$$

Therefore, $P(A=0|D=0) = \boxed{\frac{5}{6}}$

Additional Slides

Calculating Partition Function

A	B	C	D	$f_1(A=a)f_2(A=a,B=b)f_3(A=a,C=c)f_4(A=a,C=c,D=d)$
0	0	0	0	$9*6*8*1=432$
0	0	0	1	$9*6*8*9=3888$
0	0	1	0	$9*6*2*4=432$
0	0	1	1	$9*6*2*6=648$
0	1	0	0	$9*4*8*1=288$
0	1	0	1	$9*4*8*9=2592$
0	1	1	0	$9*4*2*4=288$
0	1	1	1	$9*4*2*6=432$

Calculating Partition Function

A	B	C	D	$f_1(A=a)f_2(A=a,B=b)f_3(A=a,C=c)f_4(A=a,C=c,D=d)$
1	0	0	0	$1*3*5*2=30$
1	0	0	1	$1*3*5*8=120$
1	0	1	0	$1*3*5*7=105$
1	0	1	1	$1*3*5*3=45$
1	1	0	0	$1*7*5*2=70$
1	1	0	1	$1*7*5*8=280$
1	1	1	0	$1*7*5*7=245$
1	1	1	1	$1*7*5*3=105$