50.007
Machine Learning

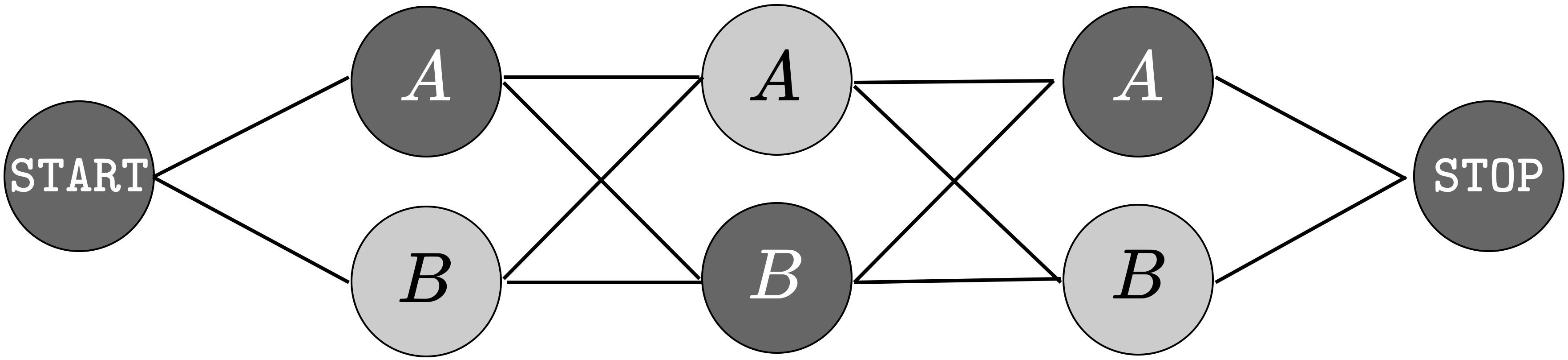
Lu, Wei

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Hidden Markov Model (IV)

Hidden Markov Model
Unsupervised Learning

We don't know the model parameters,
but only know there are two possible states: A, B.



$$\mathbf{x} = \text{the}, \text{dog}, \text{the}$$

What is the most probable $\mathbf{y}$ sequence for the given $\mathbf{x}$ sequence?

Hard EM for HMM

E-Step

Run Viterbi, and then collect counts from each instance

M-Step

$$a_{u,v} = \frac{\text{count}(u,v)}{\text{count}(u)} \qquad b_u(o) = \frac{\text{count}(u \to o)}{\text{count}(u)}$$

Soft EM for HMM

E-Step

Run forward-backward algorithm to collect fractional counts from each instance

M-Step

$$a_{u,v} = \frac{\text{count}(u,v)}{\text{count}(u)} \qquad b_u(o) = \frac{\text{count}(u \to o)}{\text{count}(u)}$$

Soft EM for HMM

$$\boxed{\text{Finding the fractional count}}$$

$$\text{count}(u, v) = \sum_{i=1}^{m} \text{count}^{(i)}(u, v)$$

Soft EM for HMM

$$\boxed{\text{Finding the fractional count}}$$

$$\text{count}(u, v) = \sum_{i=1}^{m} \text{count}^{(i)}(u, v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow v)$$

Soft EM for HMM

$$\boxed{\text{Finding the fractional count}}$$

$$\text{count}(u, v) = \sum_{i=1}^{m} \text{count}^{(i)}(u, v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \to v)$$

$$= \sum_{i=1}^{m} \boxed{\sum_{j=0}^{n} p(y_j = u, y_{j+1} = v|\mathbf{x}^{(i)})}$$

Soft EM for HMM

$$\boxed{\text{Finding the fractional count}}$$

$$\text{count}(u, v) = \sum_{i=1}^{m} \text{count}^{(i)}(u, v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)})\text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \rightarrow v)$$

$$= \sum_{i=1}^{m} \boxed{\sum_{j=0}^{n} p(y_j = u, y_{j+1} = v|\mathbf{x}^{(i)})}$$

$$\text{count}(u) = \sum_{i=1}^{m} \text{count}^{(i)}(u)$$

$$= \sum_{i=1}^{m} \boxed{\sum_{j=0}^{n} p(y_j = u|\mathbf{x}^{(i)})}$$

n here is the length of the input sentence, which may be different for a different input

Soft EM for HMM

Finding the fractional count

$$\text{count}(u,v) = \sum_{i=1}^{m} \text{count}^{(i)}(u,v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)})\text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \to v)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} p(y_j = u, y_{j+1} = v|\mathbf{x})$$

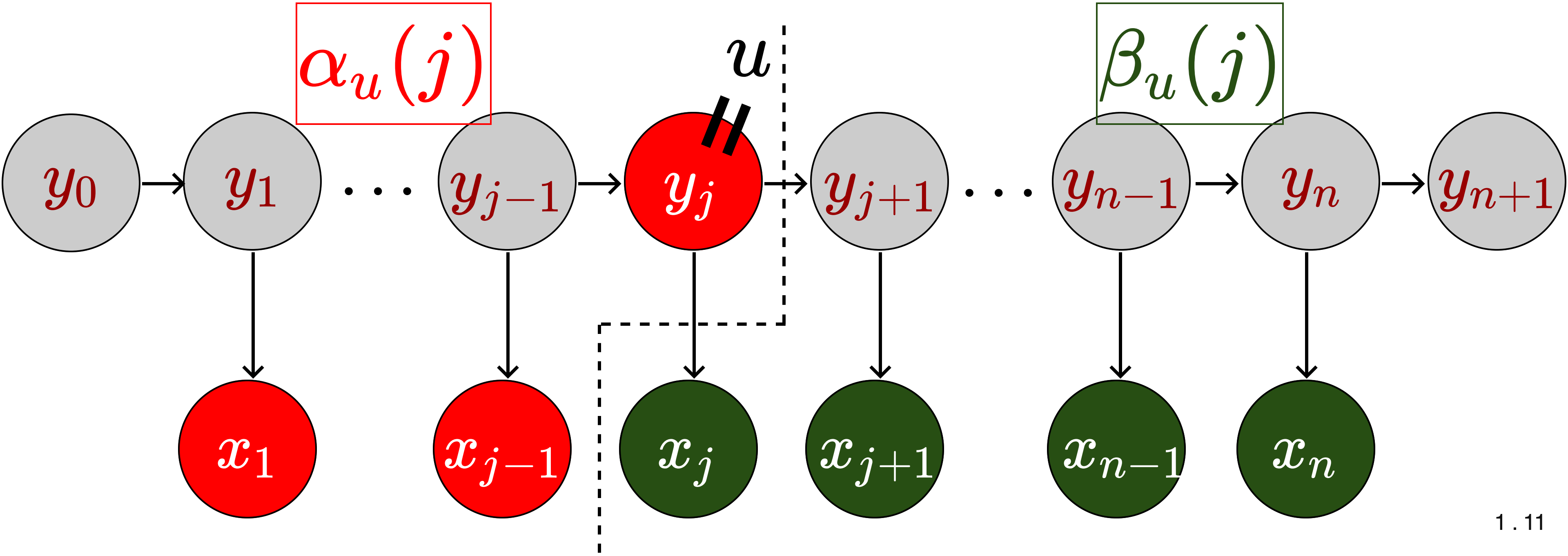
$$\text{count}(u) = \sum_{i=1}^{m} \text{count}^{(i)}(u)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} p(y_j = u|\mathbf{x})$$

Soft EM for HMM

$$\sum_{j=0}^{n} p(y_j = u|\mathbf{x})$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

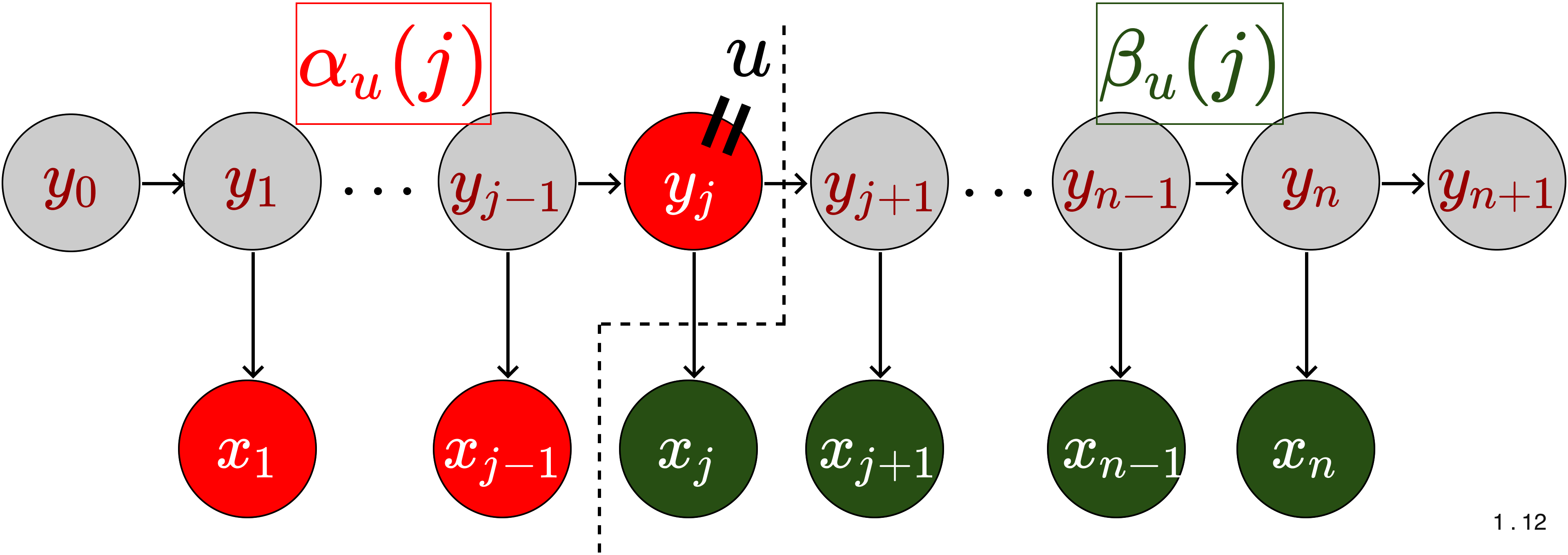


Soft EM for HMM

$$\sum_{j=0}^{n} p(y_j = u|\mathbf{x})$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

$$= \frac{\alpha_u(j)\beta_u(j)}{\quad}$$

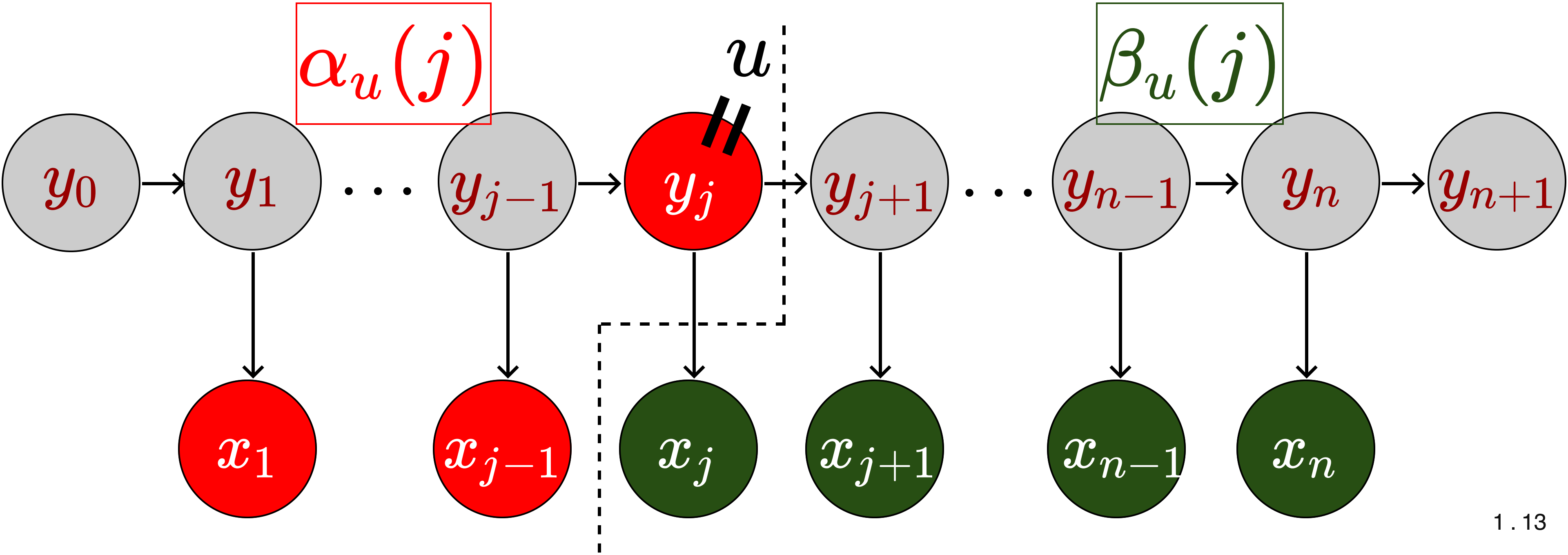


Soft EM for HMM

$$\sum_{j=0}^{n} p(y_j = u|\mathbf{x})$$

$$= \frac{p(x_1,x_2,...,x_{j-1},y_j=u,x_j,x_{j+1},...,x_n;\theta)}{\sum_v p(x_1,x_2,...,x_{k-1},y_k=v,x_k,x_{k+1},...,x_n;\theta)}$$

$$= \frac{\alpha_u(j)\beta_u(j)}{\sum_v \alpha_v(k)\beta_v(k)}$$

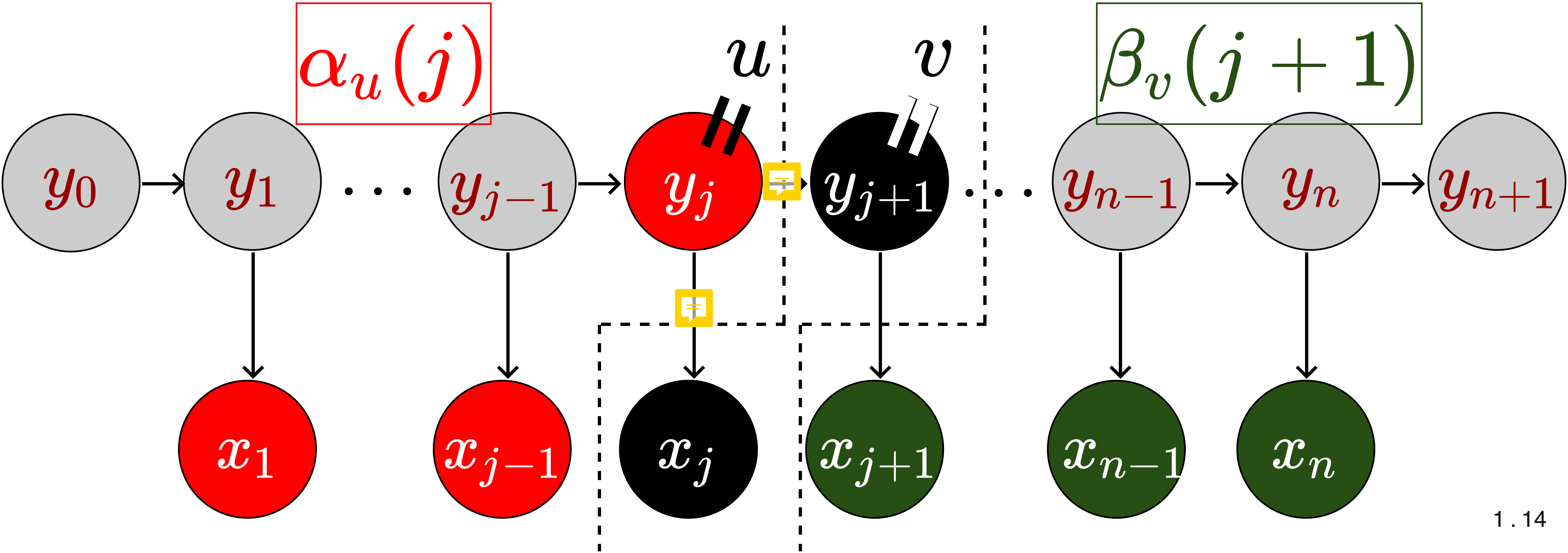


Soft EM for HMM

$$\sum_{j=0}^{n} p(y_j = u, y_{j+1} = v | \mathbf{x})$$

$$= \frac{p(x_1, x_2, \ldots, x_{j-1}, y_j = u, x_j, y_{j+1} = v, x_{j+1}, \ldots, x_n; \theta)}{\sum_v p(x_1, x_2, \ldots, x_{k-1}, y_k = v, x_k, x_{k+1}, \ldots, x_n; \theta)}$$

$$= \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k) \beta_v(k)}$$



Soft EM for HMM

Finding the fractional count

$$\text{count}(u, v) = \sum_{i=1}^{m} \text{count}^{(i)}(u, v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)}) \text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \to v)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} p(y_j = u, y_{j+1} = v|\mathbf{x})$$

$$\text{count}(u) = \sum_{i=1}^{m} \text{count}^{(i)}(u)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} p(y_j = u|\mathbf{x})$$

Soft EM for HMM

Finding the fractional count

$$\text{count}(u,v) = \sum_{i=1}^{m} \text{count}^{(i)}(u,v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)})\,\text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \to v)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k)\beta_v(k)}$$

$$\text{count}(u) = \sum_{i=1}^{m} \text{count}^{(i)}(u)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} \frac{\alpha_u(j)\beta_u(j)}{\sum_v \alpha_v(k)\beta_v(k)}$$

Soft EM for HMM

Finding the fractional count

$$\text{count}(u,v) = \sum_{i=1}^{m} \text{count}^{(i)}(u,v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)})\text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \to v)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k)\beta_v(k)}$$

$$\text{count}(u) = \sum_{i=1}^{m} \text{count}^{(i)}(u)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} \frac{\alpha_u(j)\beta_u(j)}{\sum_v \alpha_v(k)\beta_v(k)}$$

Soft EM for HMM

$$\boxed{\text{Finding the fractional count}}$$

$$\text{count}(u,v) = \sum_{i=1}^{m} \text{count}^{(i)}(u,v)$$

$$= \sum_{i=1}^{m} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}^{(i)})\text{count}(\mathbf{x}^{(i)}, \mathbf{y}, u \to v)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} \frac{\alpha_u(j) \cdot b_u(x_j) \cdot a_{u,v} \cdot \beta_v(j+1)}{\sum_v \alpha_v(k)\beta_v(k)}$$

$$\text{count}(u) = \sum_{i=1}^{m} \text{count}^{(i)}(u)$$

$$= \sum_{i=1}^{m} \sum_{j=0}^{n} \frac{\alpha_u(j)\beta_u(j)}{\sum_v \alpha_v(k)\beta_v(k)}$$

In the M-Step:

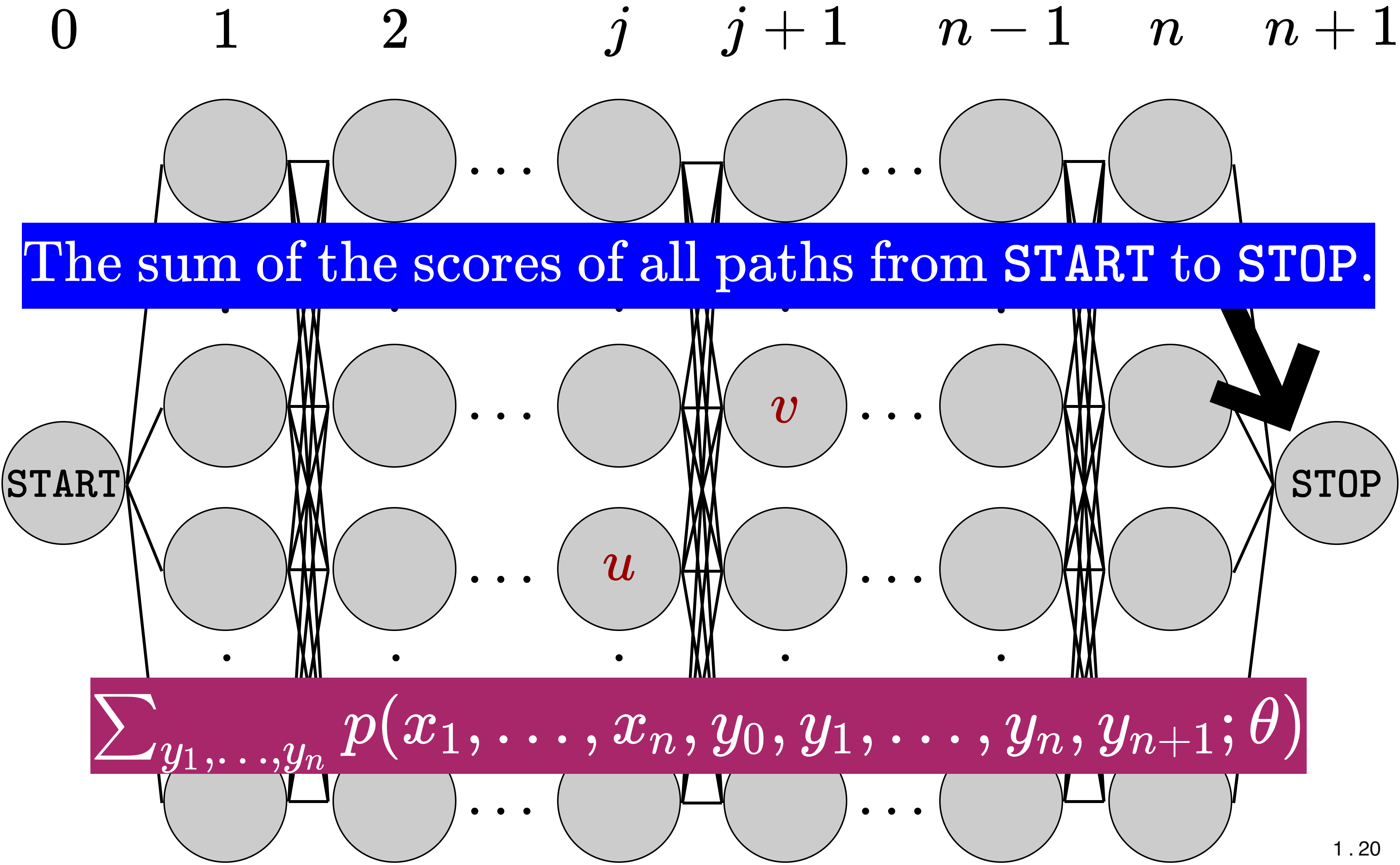
$$a_{u,v} = \frac{\text{count}(u,v)}{\text{count}(u)}$$

Question

How to find an efficient procedure to calculate forward and backward probabilities?

Calculate forward/backward scores efficiently → Perform inference efficiently → Calculate the expected counts efficiently → Perform the soft-EM efficiently

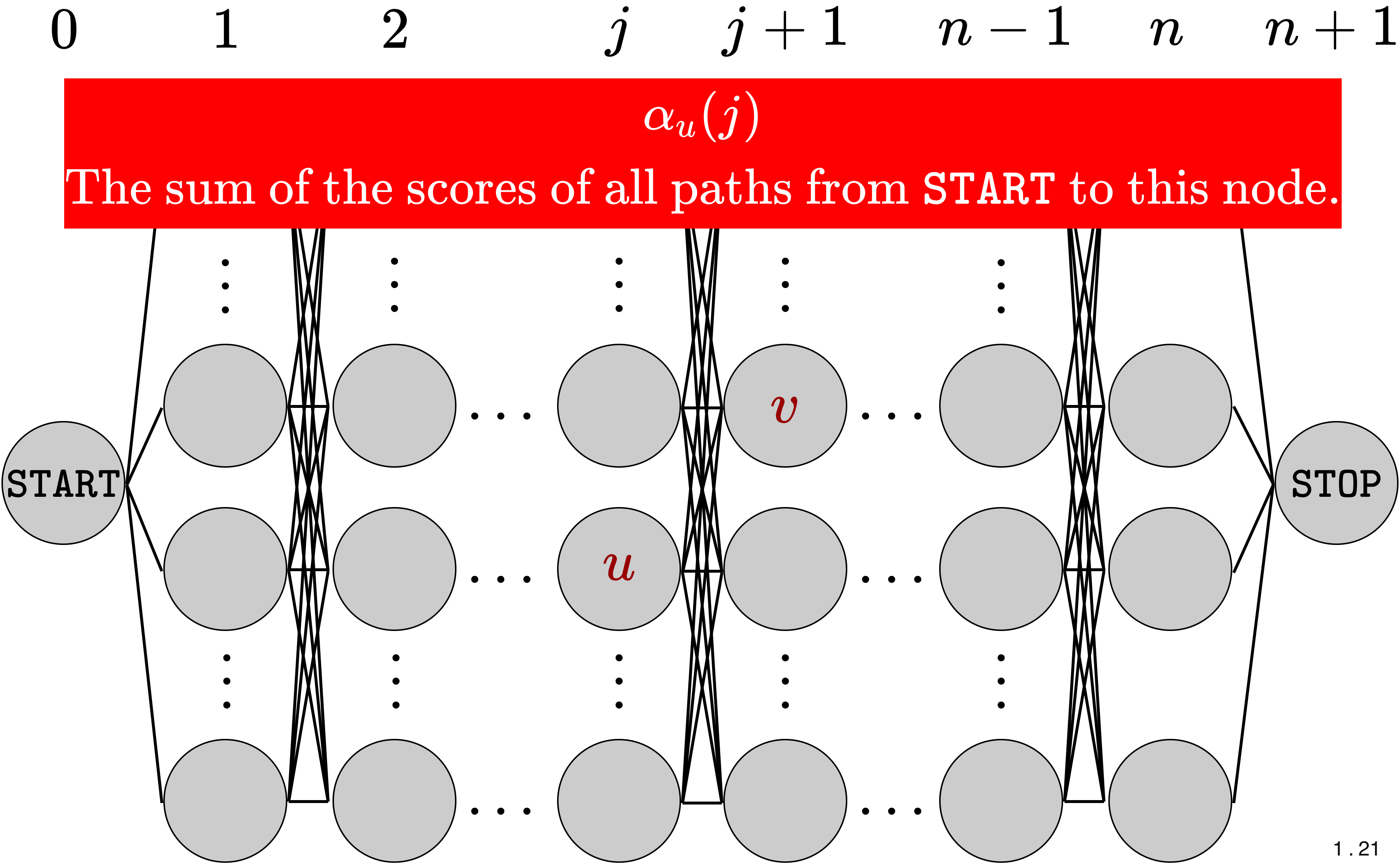
Inference in HMM



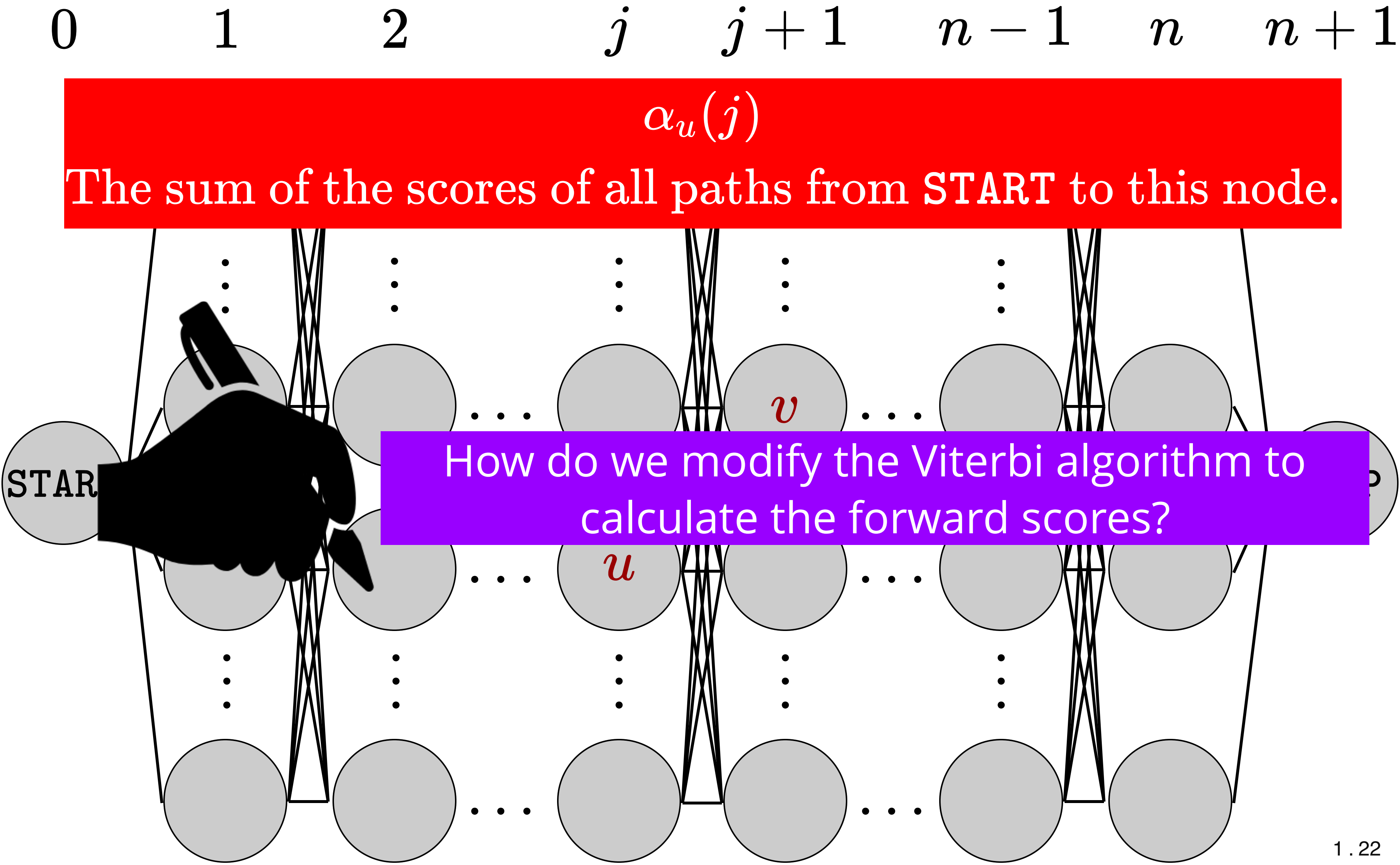
The sum of the scores of all paths from **START** to **STOP**.

$$\sum_{y_1,\ldots,y_n} p(x_1,\ldots,x_n,y_0,y_1,\ldots,y_n,y_{n+1};\theta)$$

Forward-Backward Algorithm



Forward-Backward Algorithm

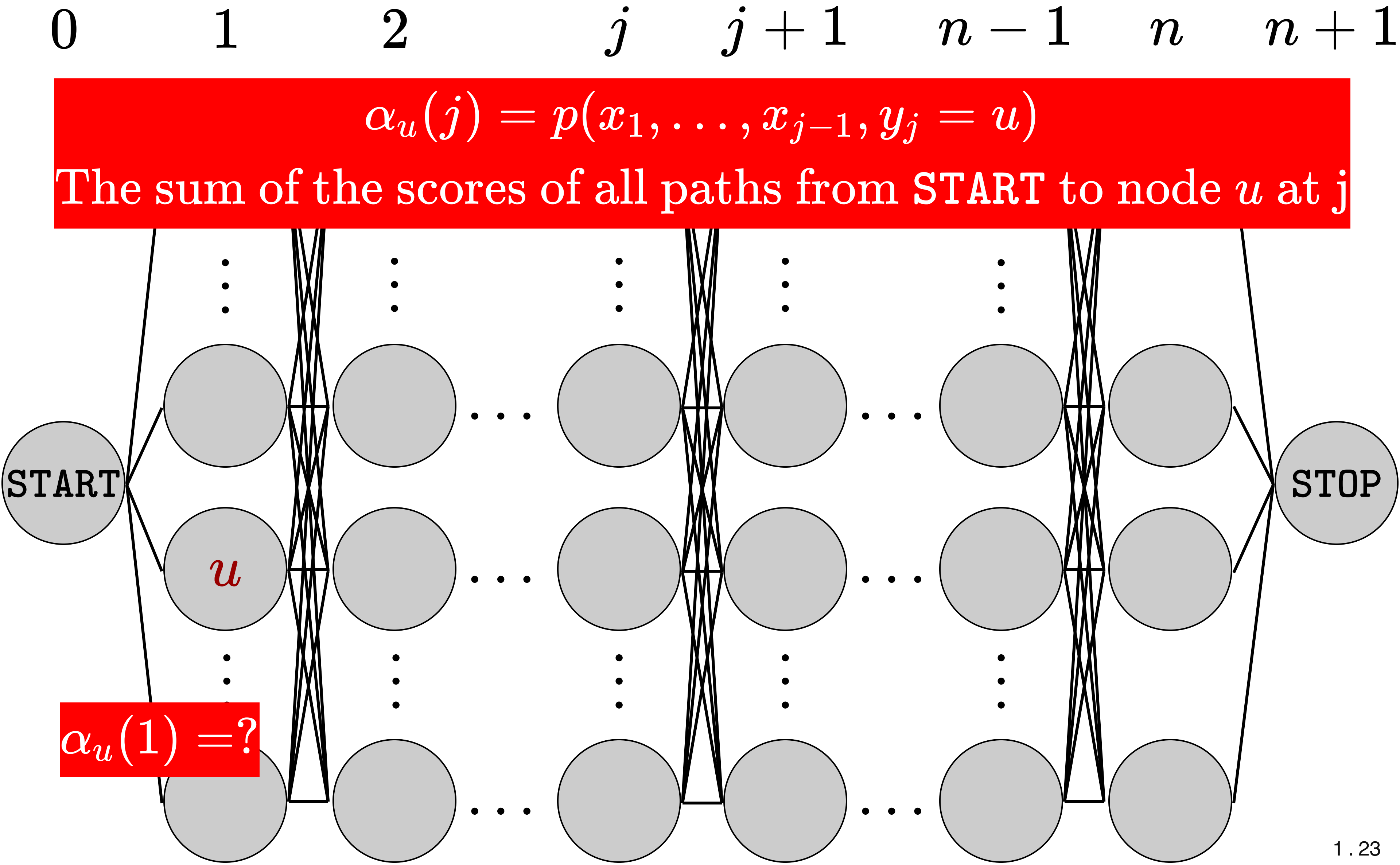


$\alpha_u(j)$

The sum of the scores of all paths from **START** to this node.

How do we modify the Viterbi algorithm to calculate the forward scores?

Forward-Backward Algorithm

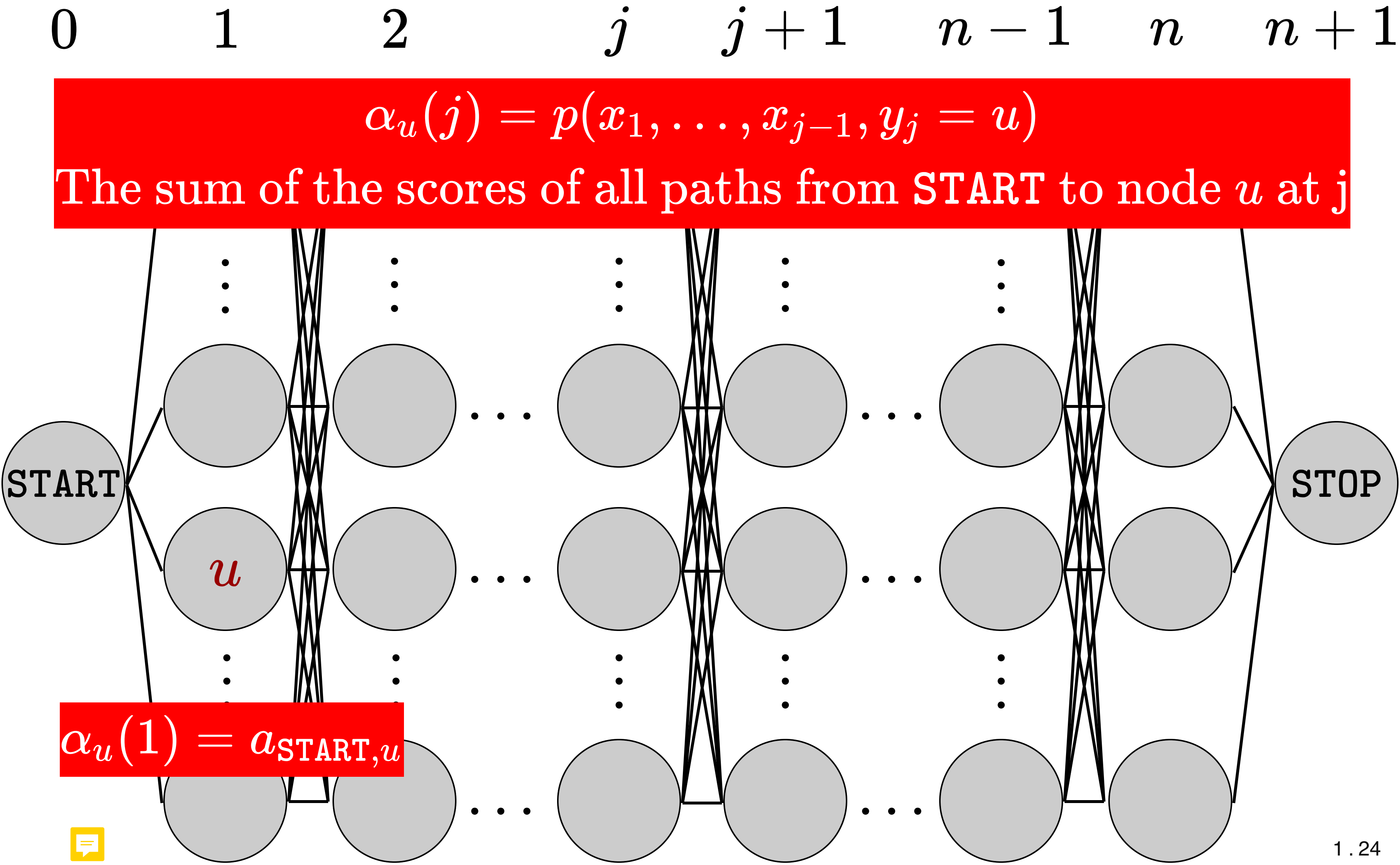


$$\alpha_u(j) = p(x_1, \ldots, x_{j-1}, y_j = u)$$

The sum of the scores of all paths from **START** to node u at j

$\alpha_u(1) = ?$

Forward-Backward Algorithm

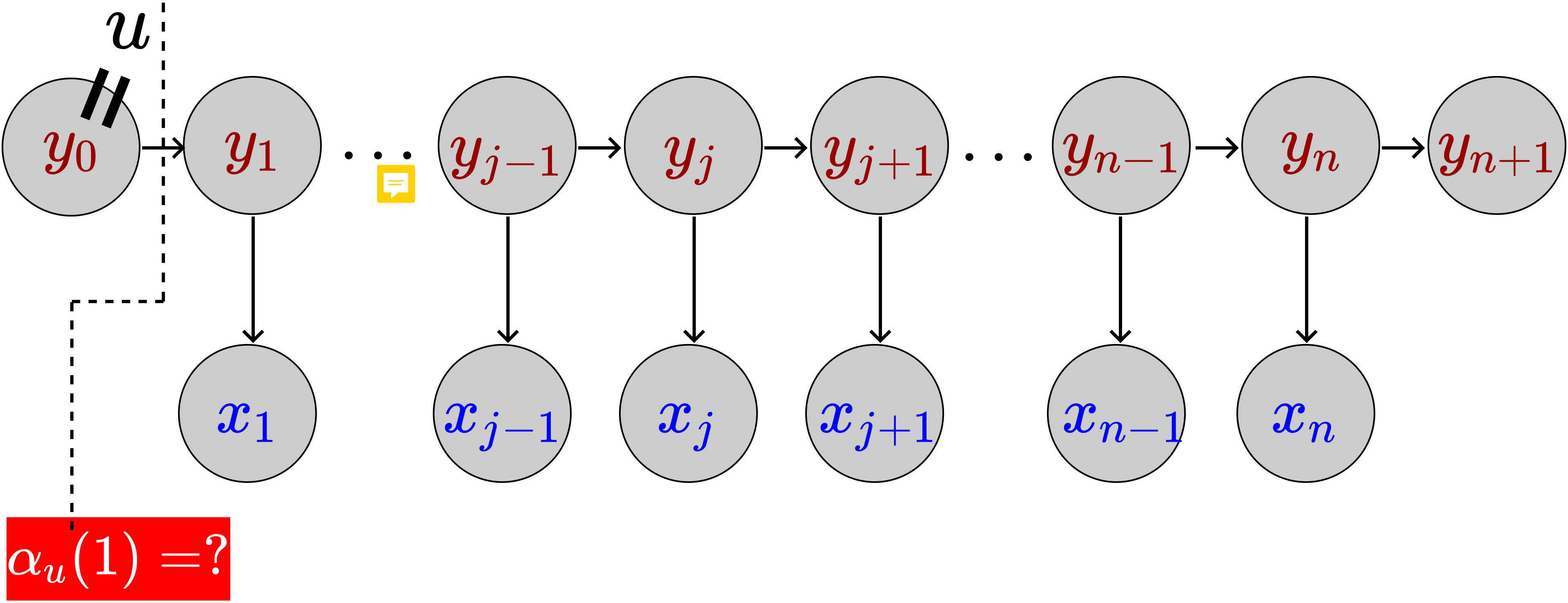


$$\alpha_u(j) = p(x_1, \dots, x_{j-1}, y_j = u)$$

The sum of the scores of all paths from **START** to node u at j

$$\alpha_u(1) = a_{\mathtt{START},u}$$

Forward-Backward Algorithm

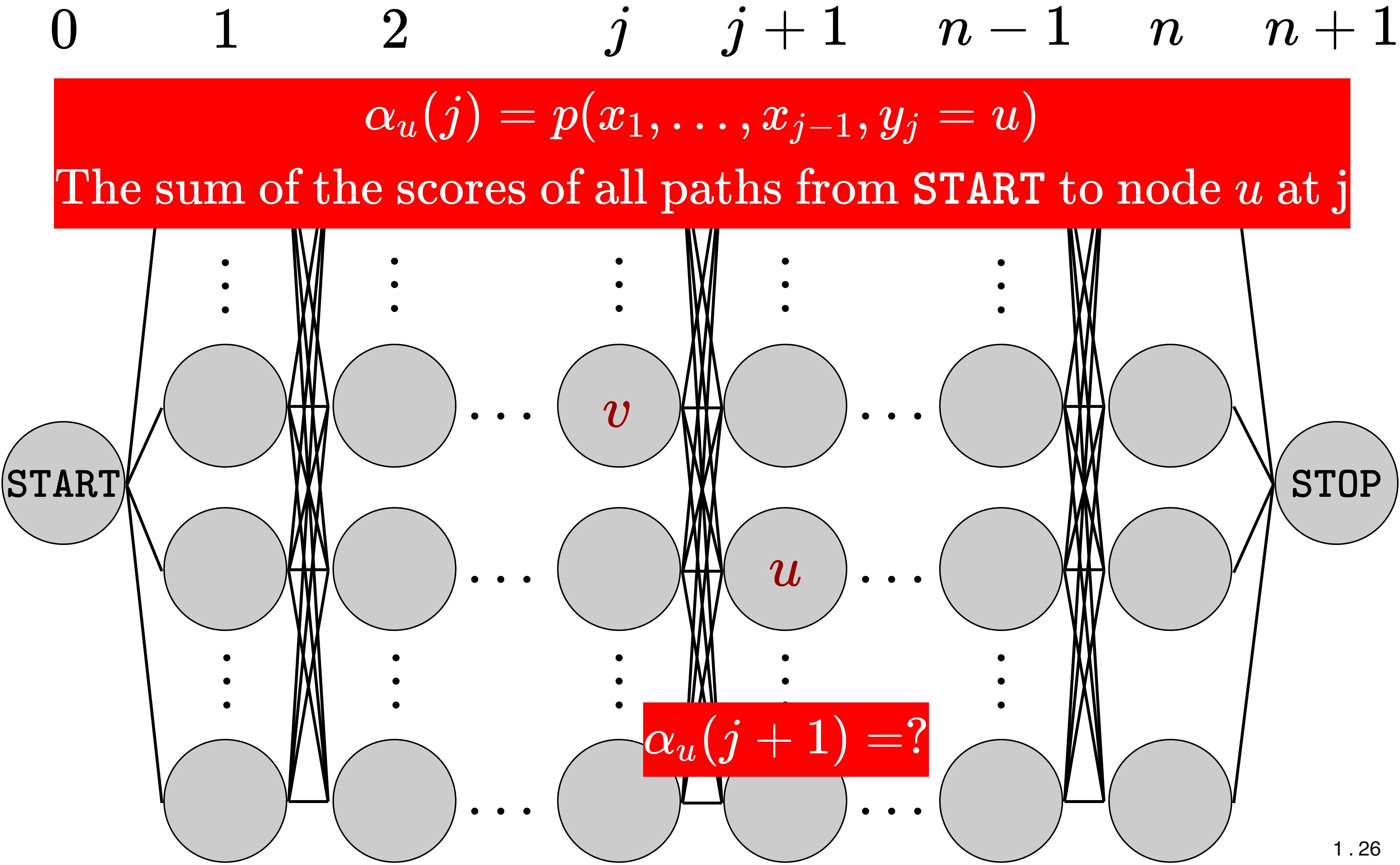


$$\boxed{\alpha_u(1) = a_{\texttt{START},u}}$$

Forward-Backward Algorithm

$$\alpha_u(j) = p(x_1, \ldots, x_{j-1}, y_j = u)$$

The sum of the scores of all paths from **START** to node u at j

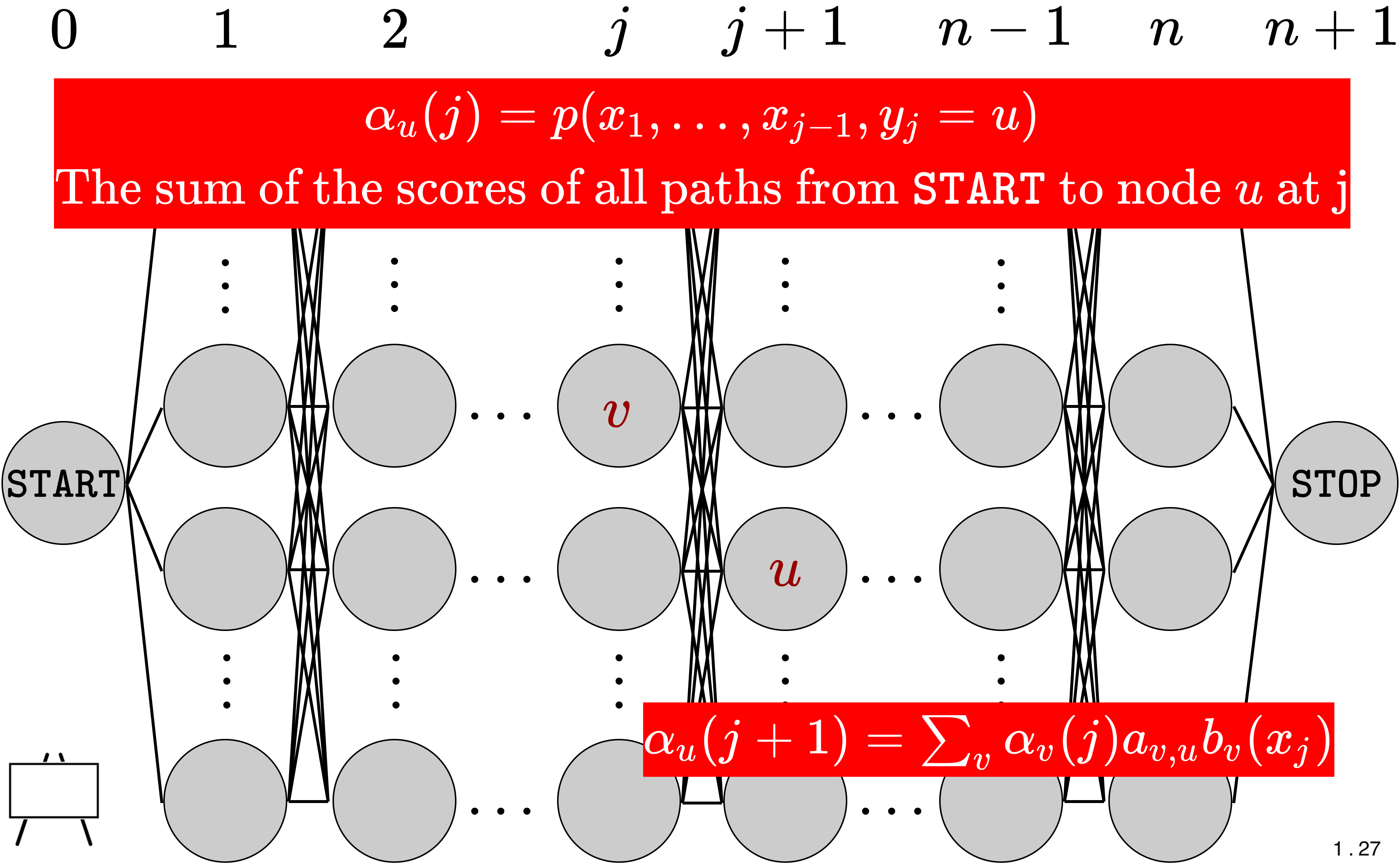


$$\alpha_u(j+1) = ?$$

Forward-Backward Algorithm

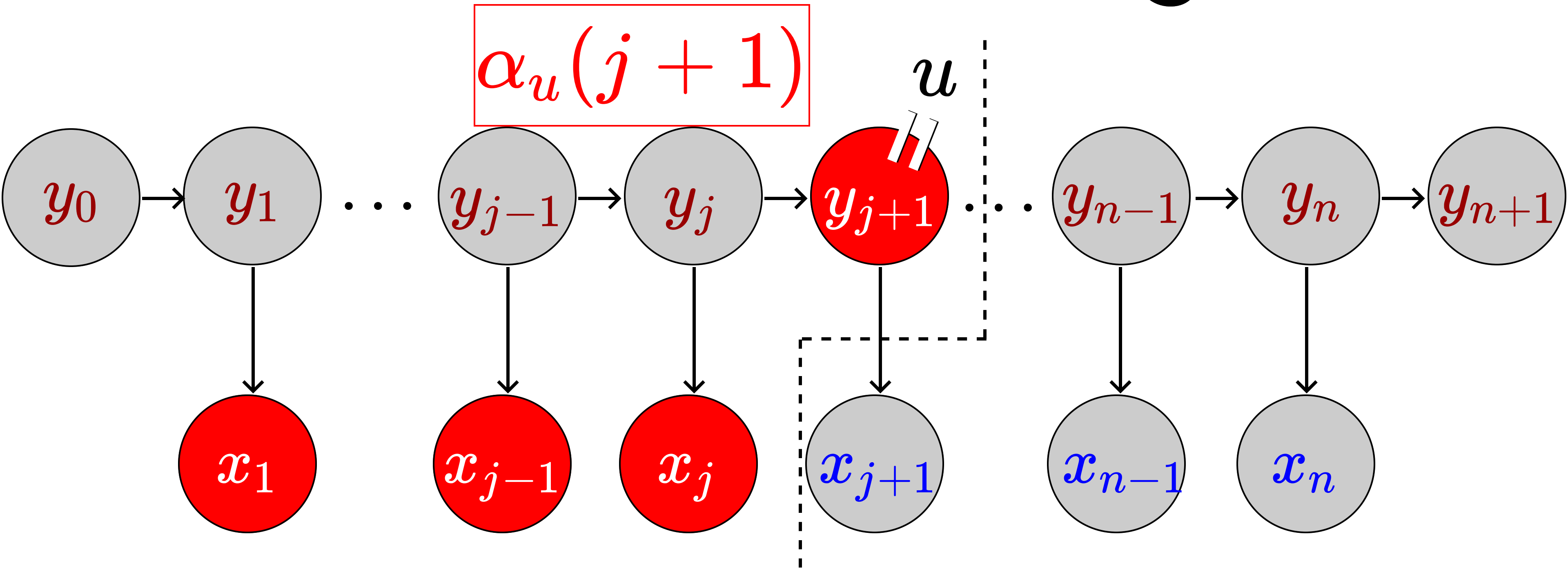
$$\alpha_u(j) = p(x_1, \ldots, x_{j-1}, y_j = u)$$

The sum of the scores of all paths from **START** to node u at j

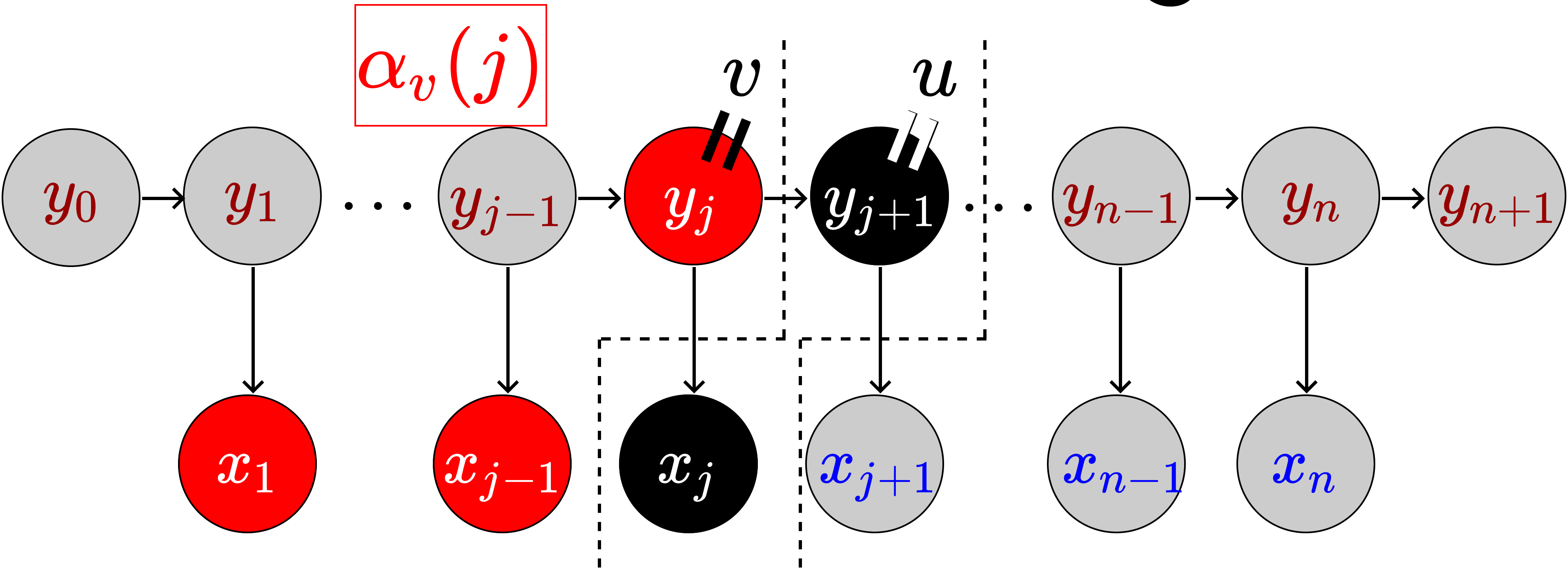


$$\alpha_u(j+1) = \sum_v \alpha_v(j) a_{v,u} b_v(x_j)$$

Forward-Backward Algorithm



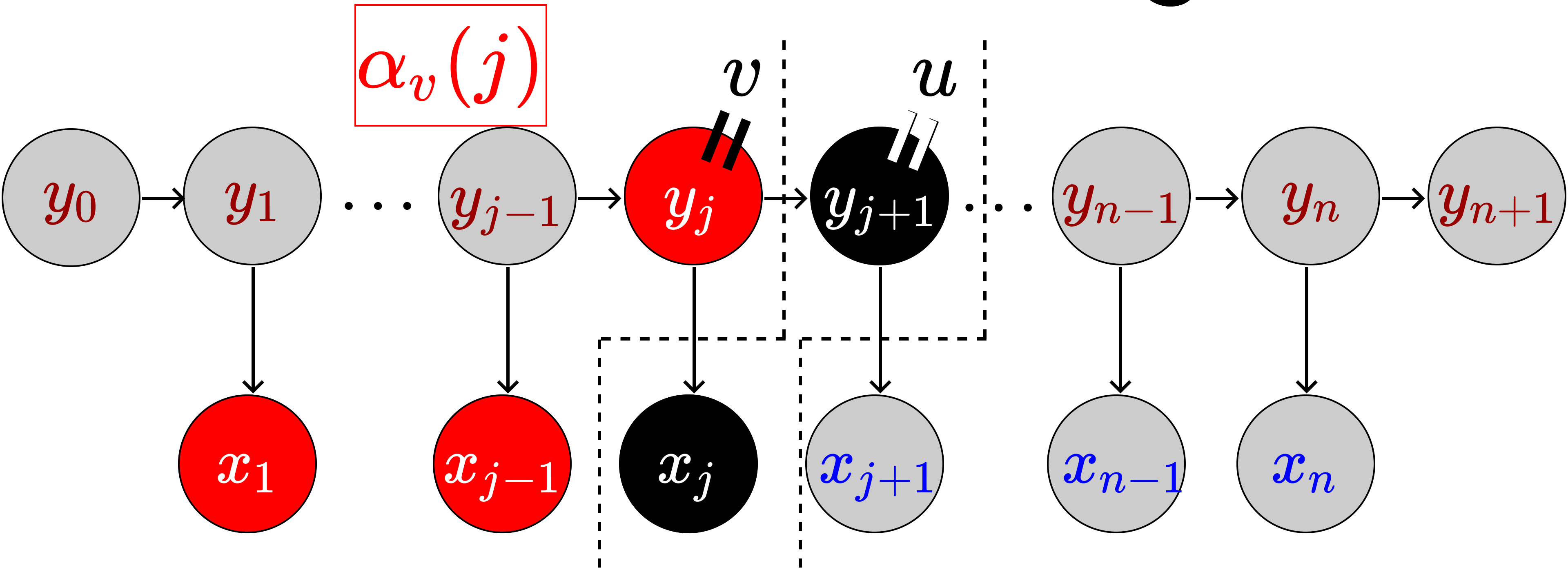
Forward-Backward Algorithm



Assuming the previous state is v

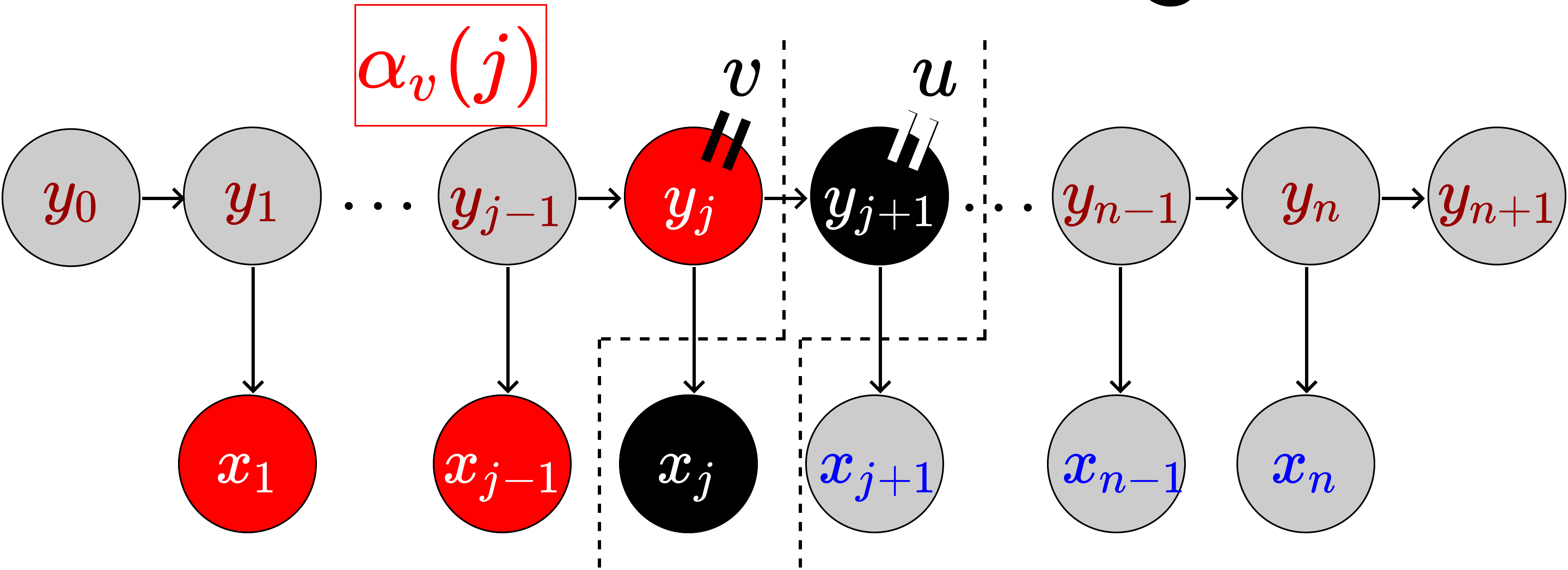
How do we generate these two nodes in black?

Forward-Backward Algorithm



$$\alpha_v(j)a_{v,u}b_v(x_j)$$

Forward-Backward Algorithm

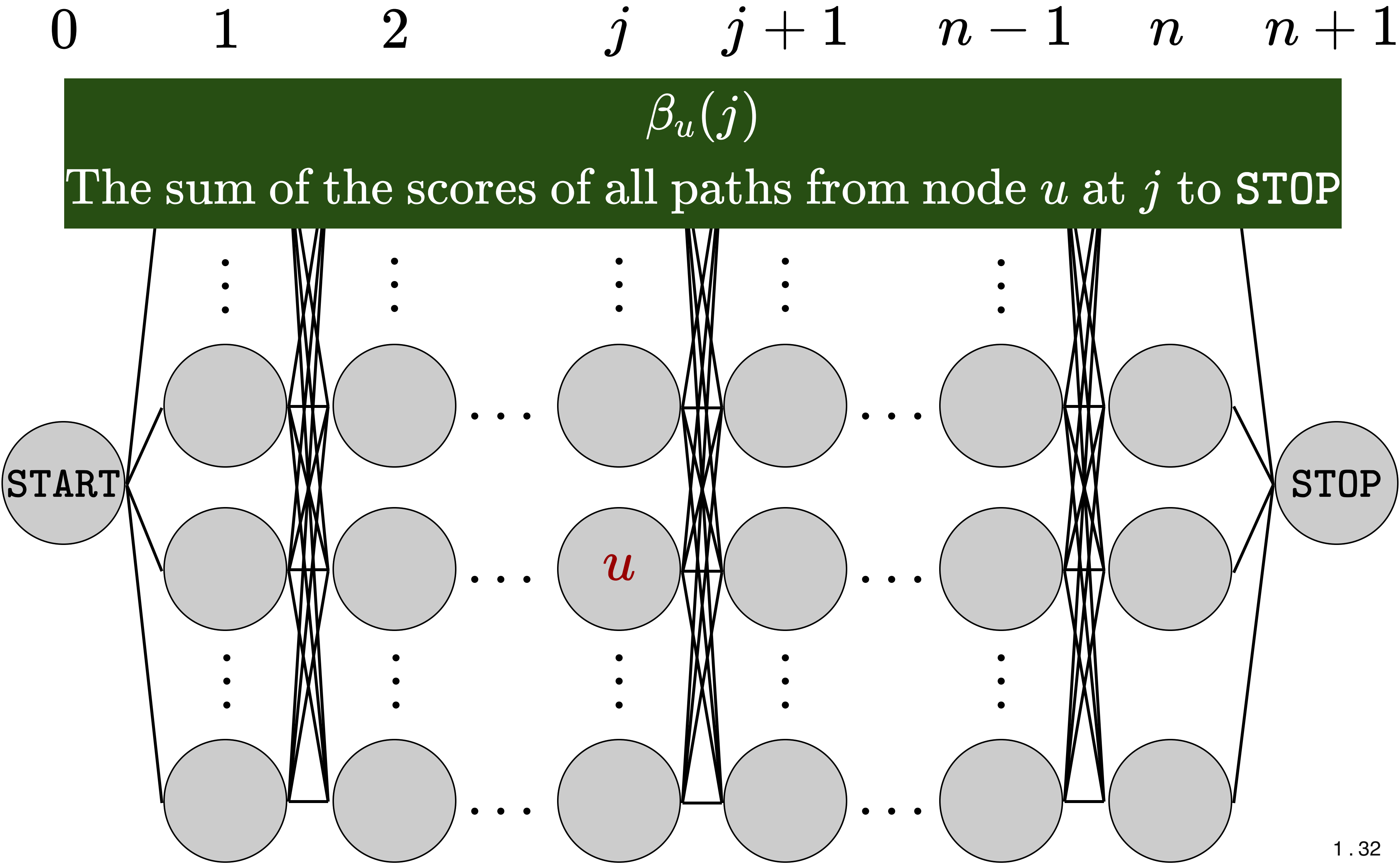


$$\boxed{\alpha_u(j+1) = \sum_v \alpha_v(j) a_{v,u} b_v(x_j)}$$

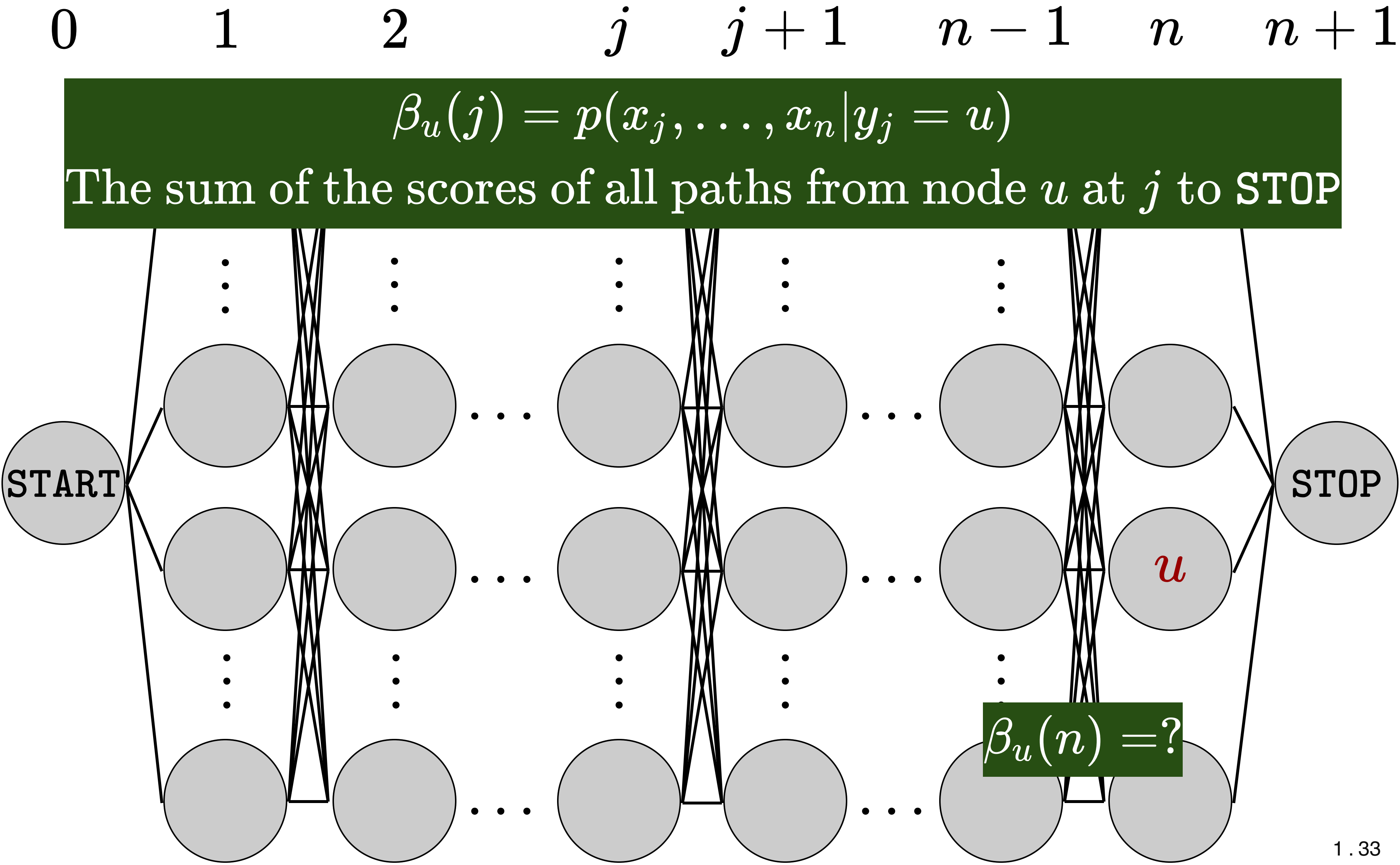
Forward-Backward Algorithm

$\beta_u(j)$

The sum of the scores of all paths from node u at j to STOP



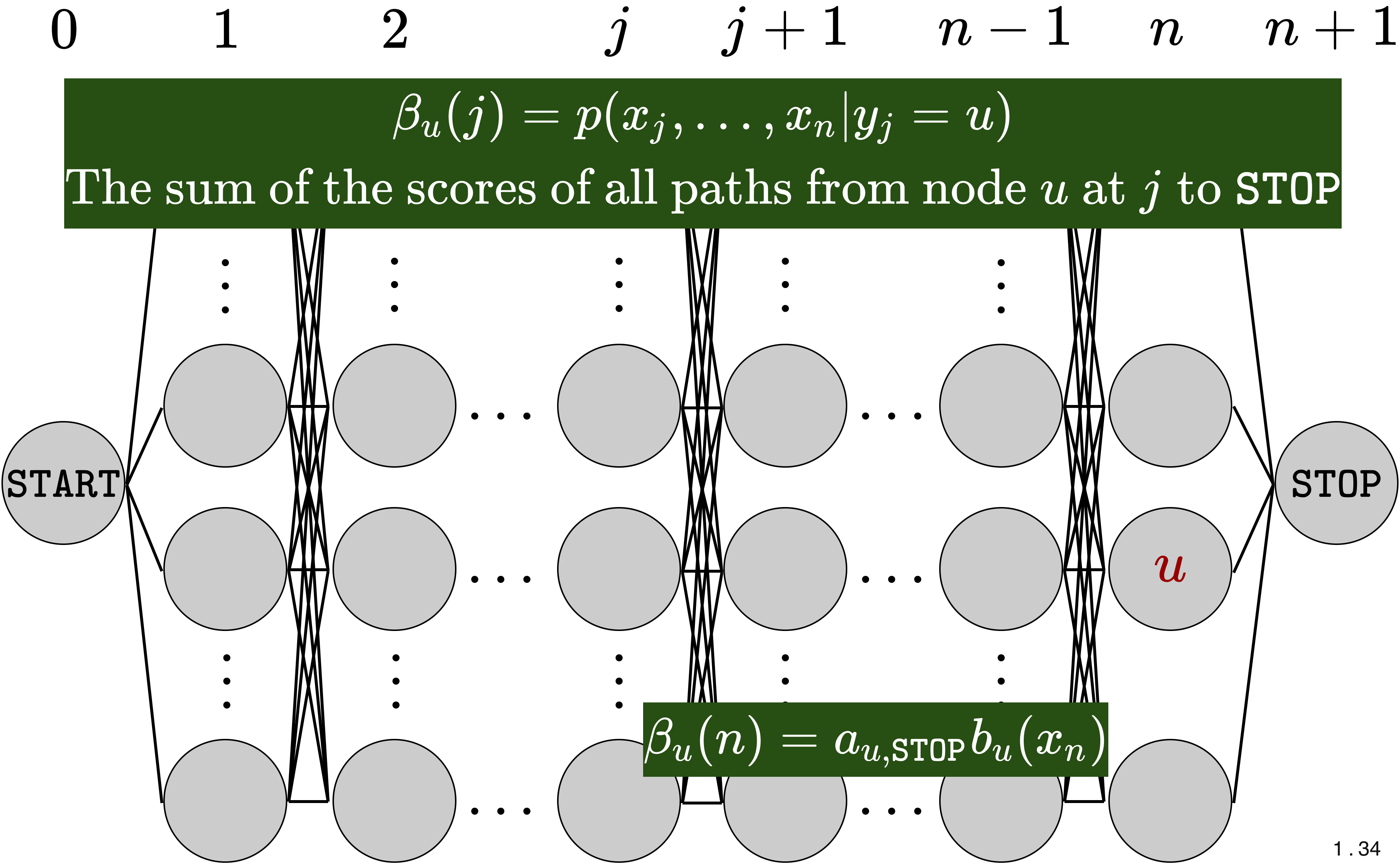
Forward-Backward Algorithm



Forward-Backward Algorithm

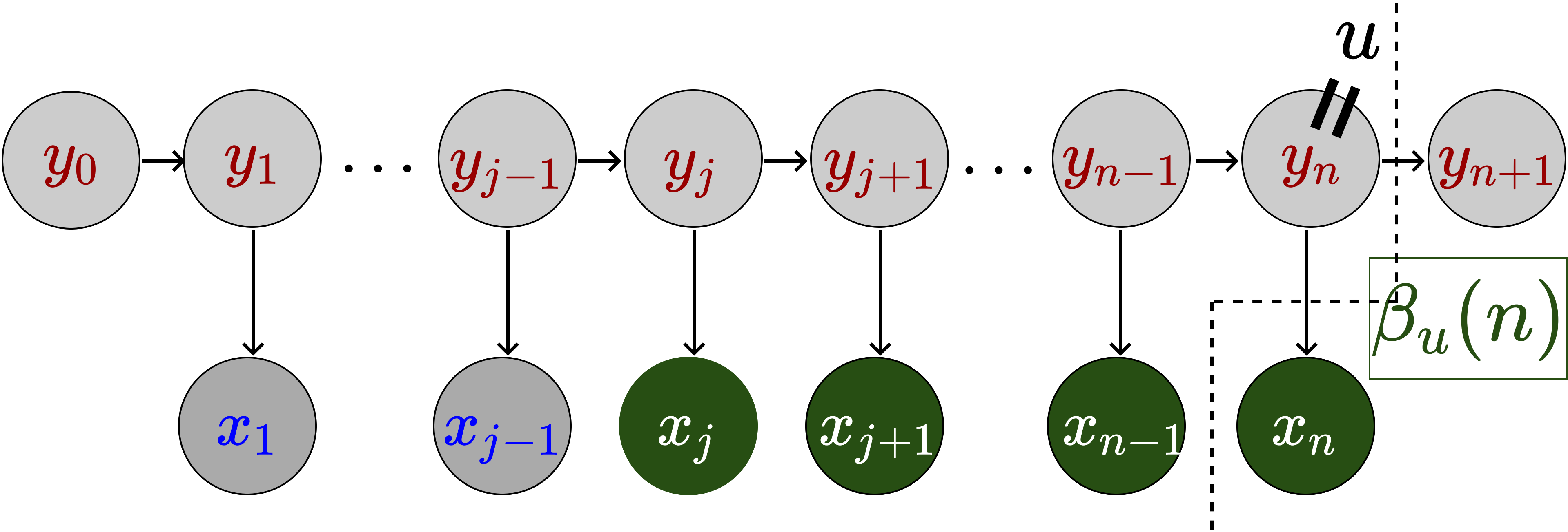
$$\beta_u(j) = p(x_j, \ldots, x_n | y_j = u)$$

The sum of the scores of all paths from node u at j to STOP



$$\beta_u(n) = a_{u,\texttt{STOP}} b_u(x_n)$$

Forward-Backward Algorithm

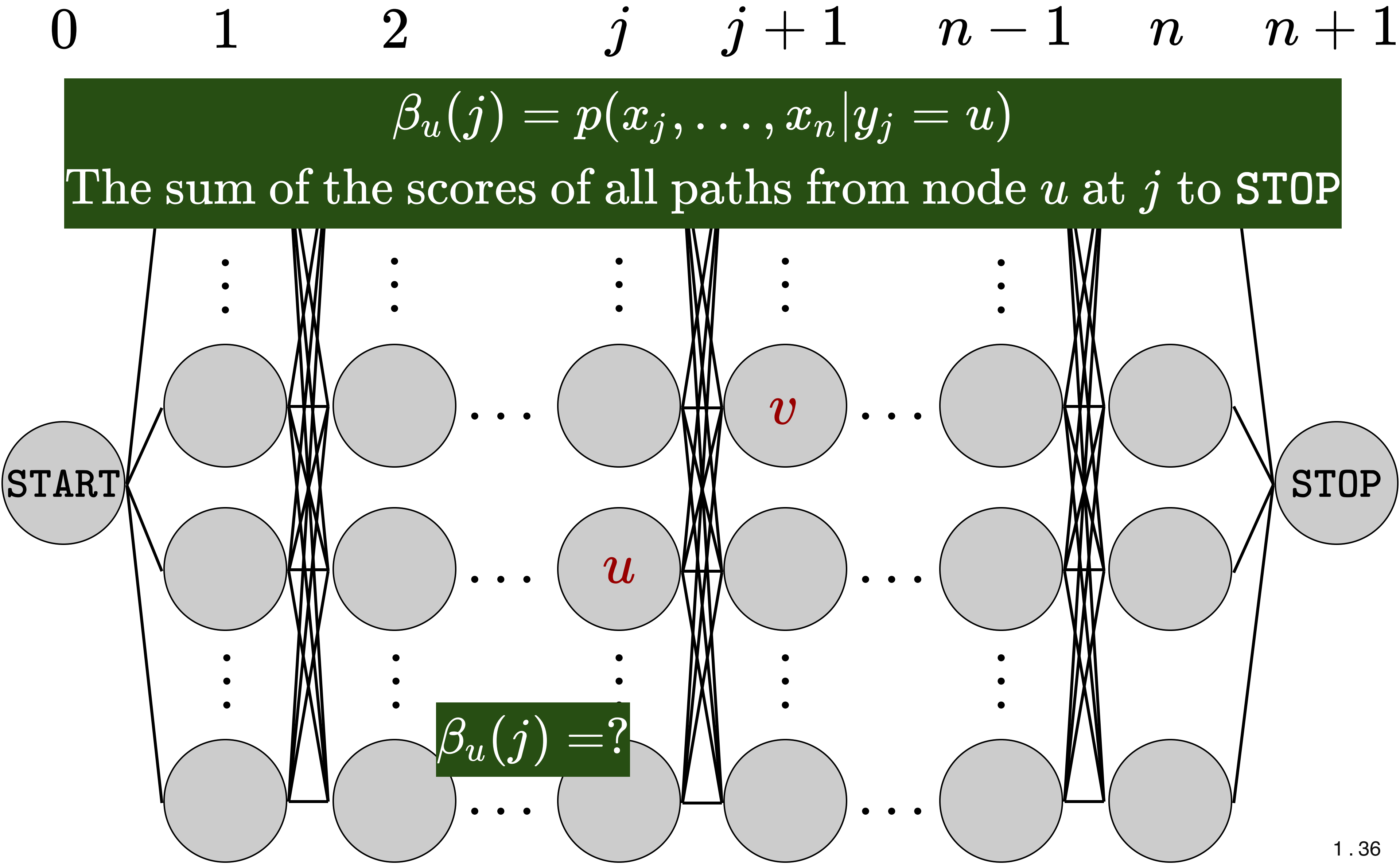


$$\boxed{\beta_u(n) = a_{u,\texttt{STOP}} b_u(x_n)}$$

Forward-Backward Algorithm

$$\beta_u(j) = p(x_j, \ldots, x_n | y_j = u)$$

The sum of the scores of all paths from node u at j to STOP

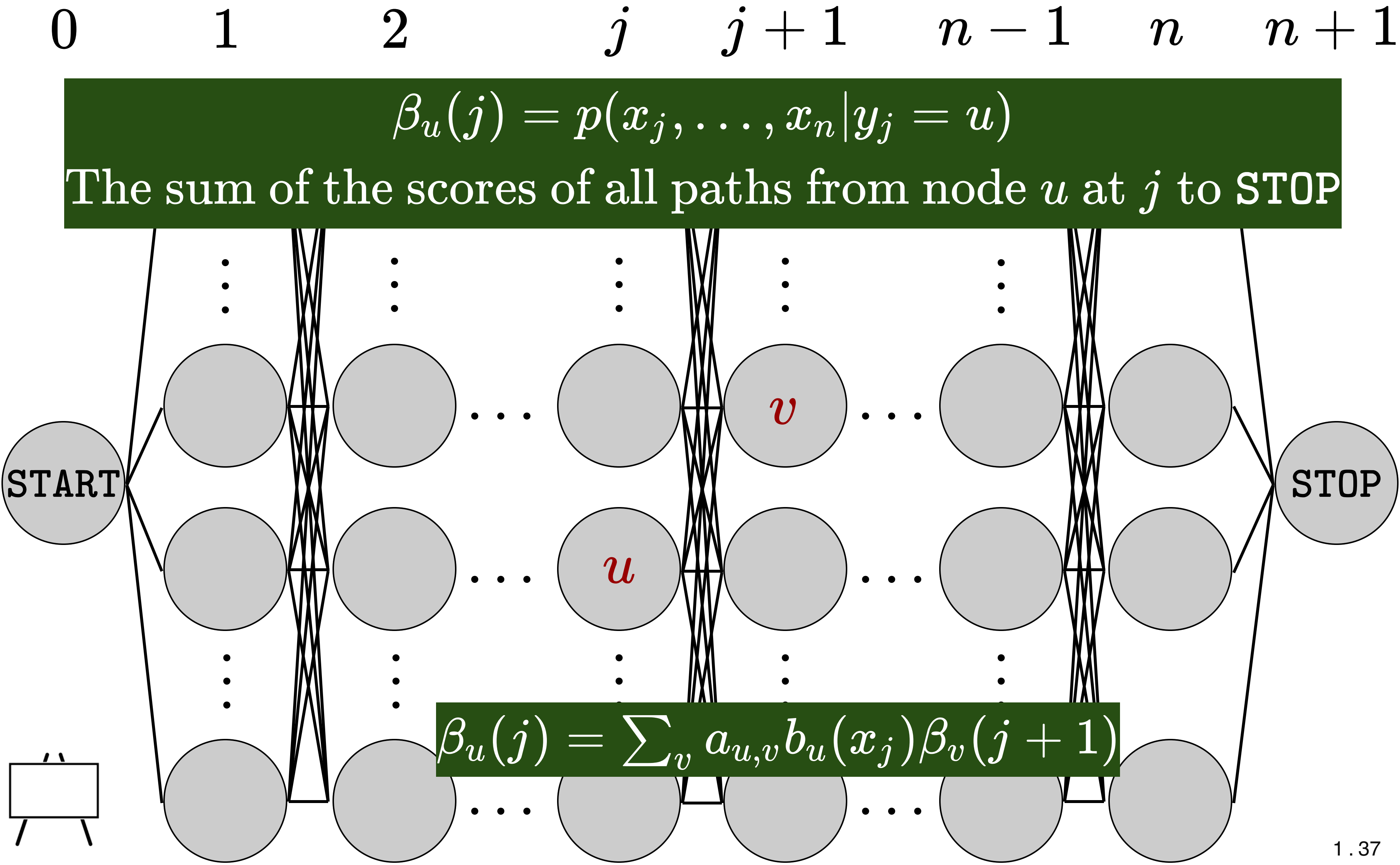


$\beta_u(j) = ?$

Forward-Backward Algorithm

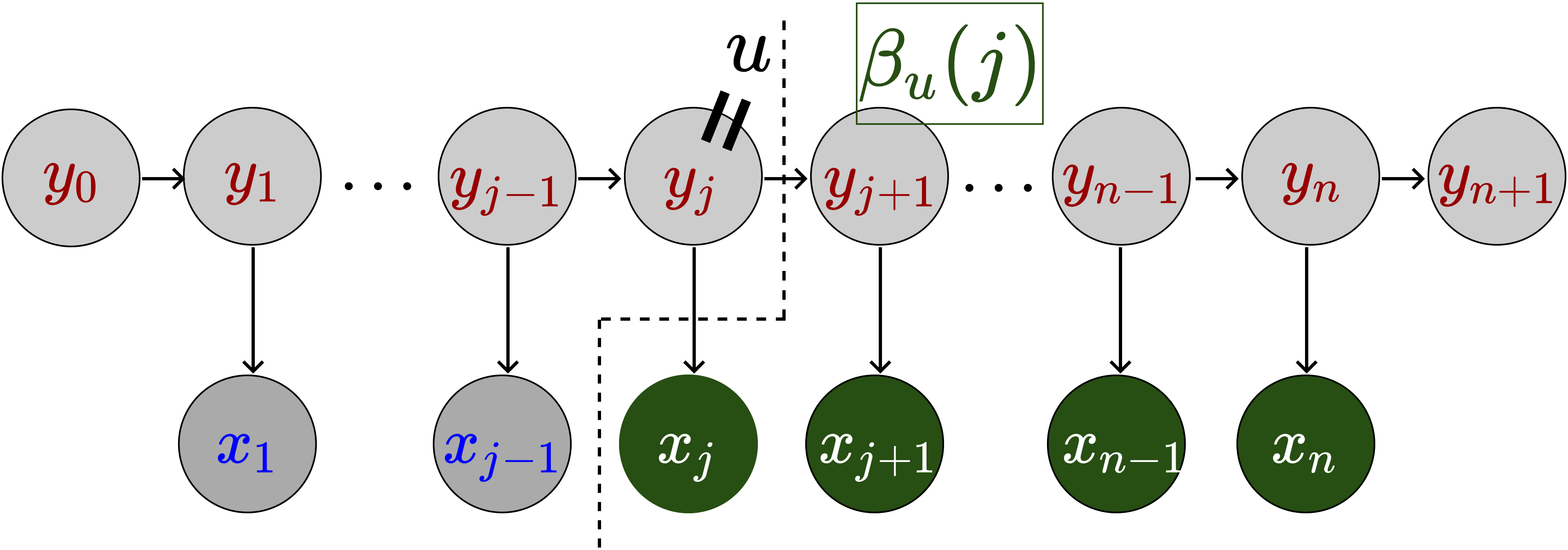
$$\beta_u(j) = p(x_j, \ldots, x_n | y_j = u)$$

The sum of the scores of all paths from node u at j to STOP

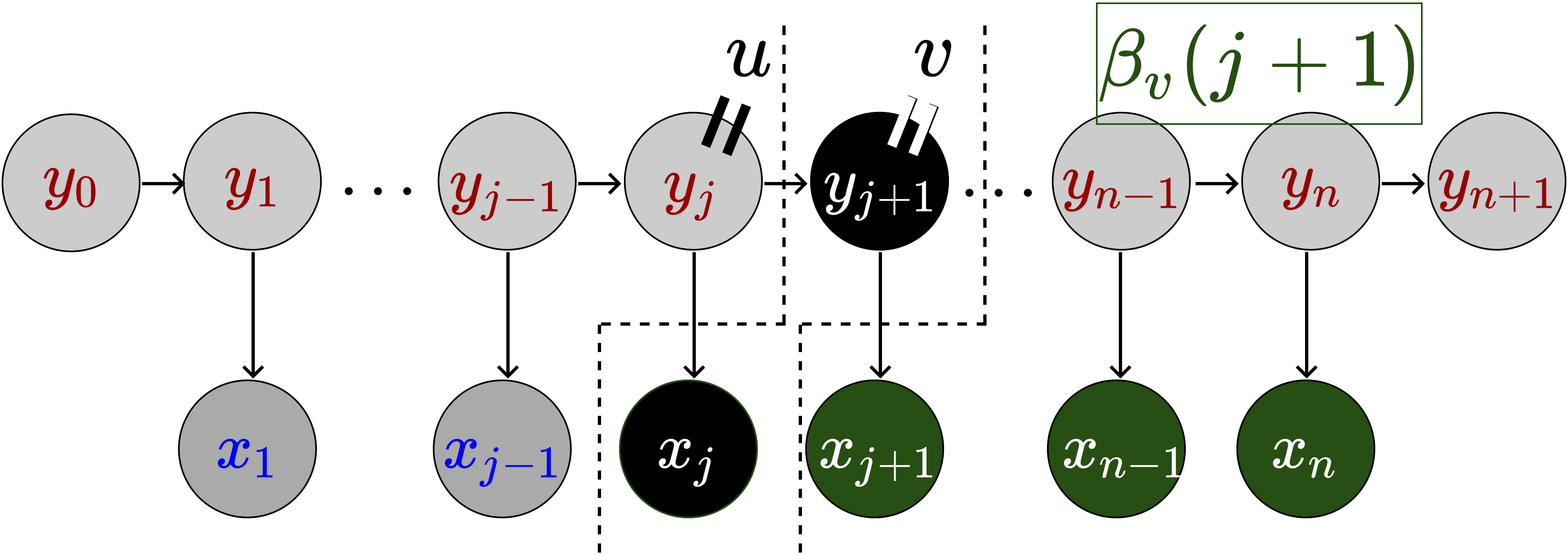


$$\beta_u(j) = \sum_v a_{u,v} b_u(x_j) \beta_v(j+1)$$

Forward-Backward Algorithm



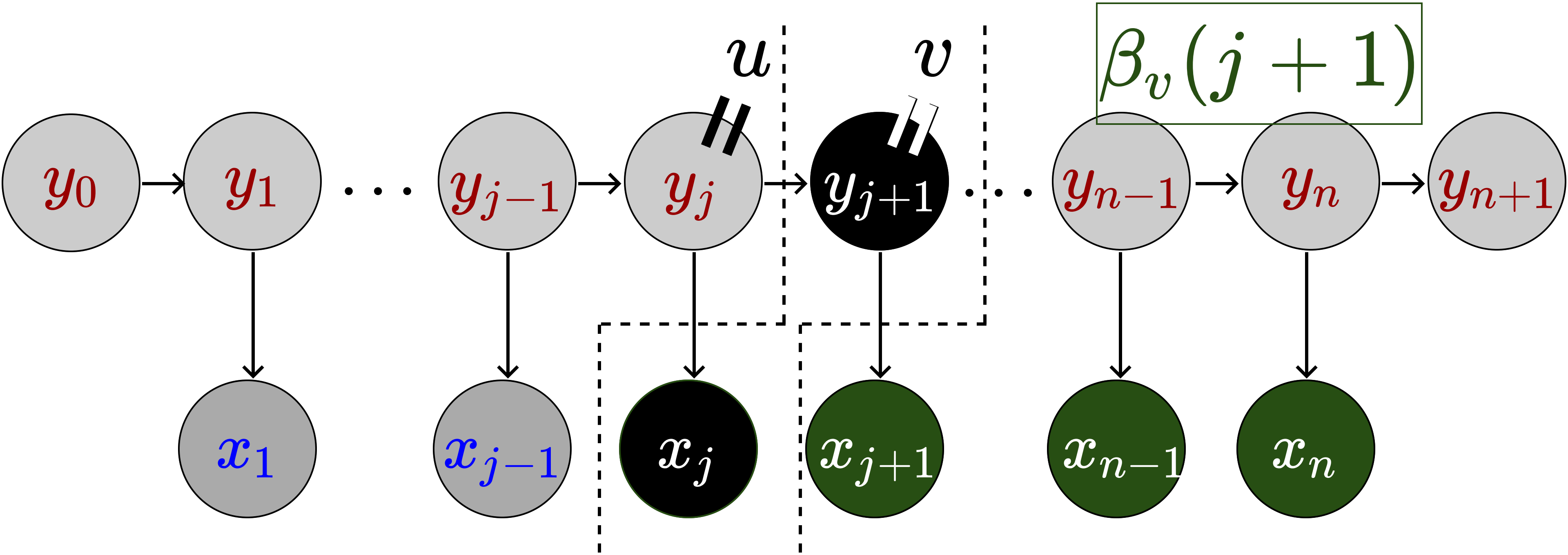
Forward-Backward Algorithm



Assuming the next state is v

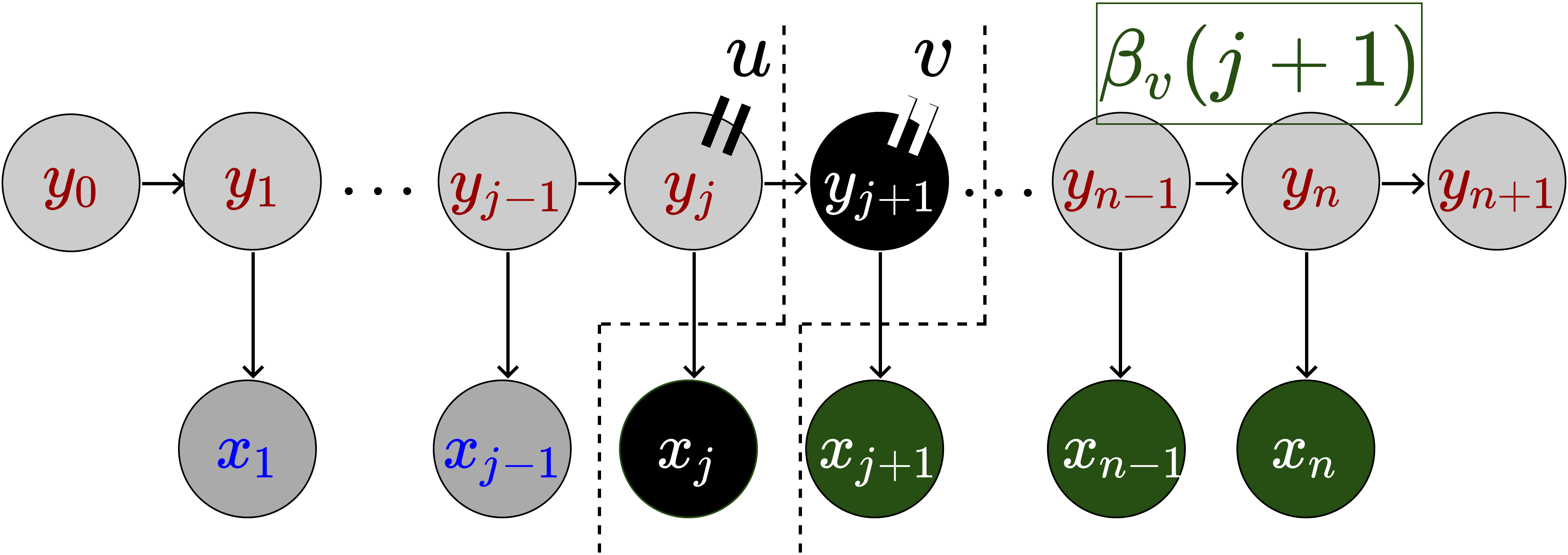
How do we generate these two nodes in black?

Forward-Backward Algorithm



$$a_{u,v} b_u(x_j) \beta_v(j+1)$$

Forward-Backward Algorithm



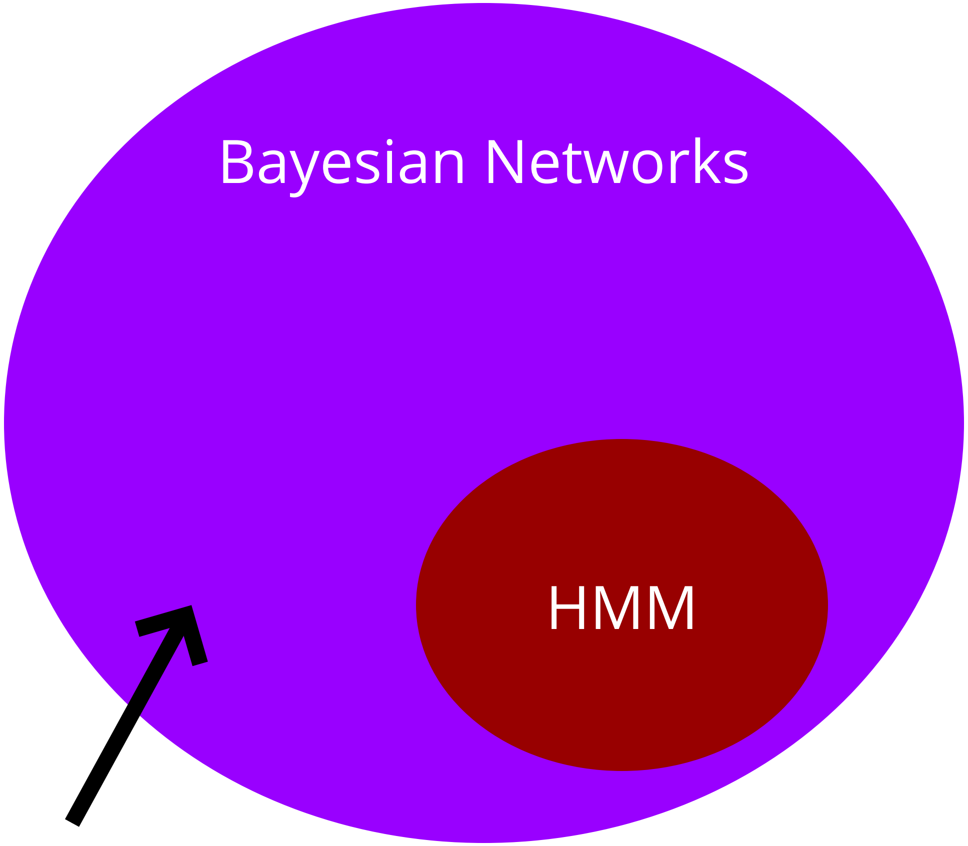
$$\boxed{\beta_u(j) = \sum_v a_{u,v} b_u(x_j) \beta_v(j+1)}$$

Question

What is the time complexity of the forward backward algorithm?

Hidden Markov Model

Bayesian Networks



Where we will be next

Subtitle