

# Genome-540 class 15

Chengxiang Qiu

# A general definition of HMM

$$HMM = (S, V, B, A, \Pi)$$

**N states**

$$S = \{s_1, \dots, s_N\}$$

**M symbols**

$$V = \{v_1, \dots, v_M\}$$

**Initial state probability:**

$$\Pi = \{\pi_1, \dots, \pi_N\} \quad \sum_{i=1}^N \pi_i = 1$$

$\pi_i$ : prob of starting at state  $s_i$

**State transition probability:**

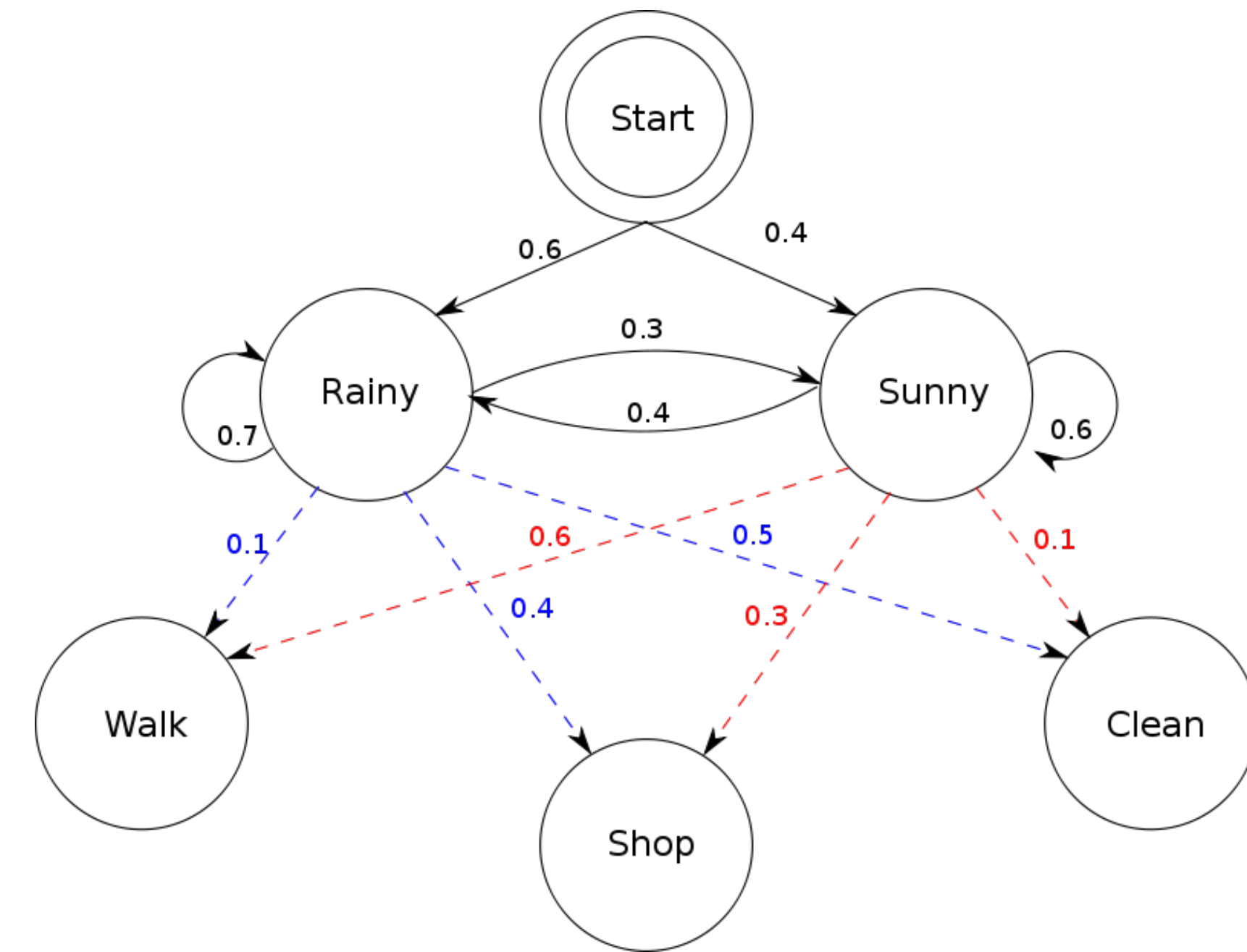
$$A = \{a_{ij}\} \quad 1 \leq i, j \leq N \quad \sum_{j=1}^N a_{ij} = 1$$

$a_{ij}$ : prob of going  $s_i \rightarrow s_j$

**Output probability:**

$$B = \{b_i(v_k)\} \quad 1 \leq i \leq N, 1 \leq k \leq M \quad \sum_{k=1}^M b_i(v_k) = 1$$

$b_i(v_k)$ : prob of "generating"  $v_k$  at  $s_i$



## COURSE-RELATED MATERIALS:

- [Math Notation](#)
- [Biological Review Slides](#): Gene and genome structure in prokaryotes and eu databases.
- [Nature paper on human genome sequence](#)
- [Nature paper on mouse genome sequence](#)
- [Siepel et al. paper on PhyloHMMs & sequence conservation](#)
- [Rabiner tutorial on HMMs](#)
- [HMM scaling tutorial \(Tobias Mann\)](#)

# Forward variables

The Forward-Backward Procedure [2], [3]<sup>b</sup>: Consider the forward variable  $\alpha_t(i)$  defined as

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i | \lambda) \quad (18)$$

i.e., the probability of the partial observation sequence,  $O_1 O_2 \cdots O_t$ , (until time  $t$ ) and state  $S_i$  at time  $t$ , given the model  $\lambda$ . We can solve for  $\alpha_t(i)$  inductively, as follows:

1) Initialization:

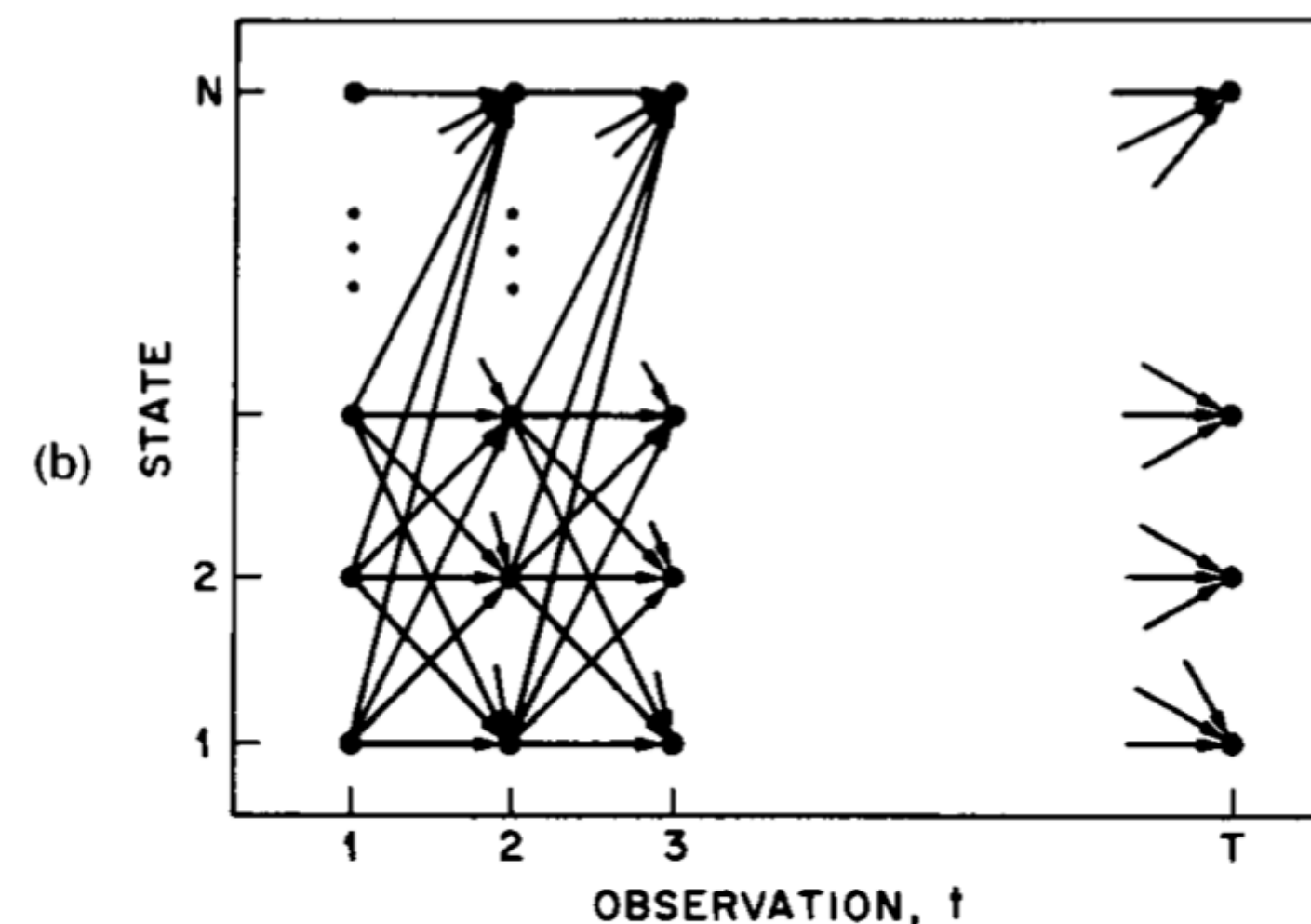
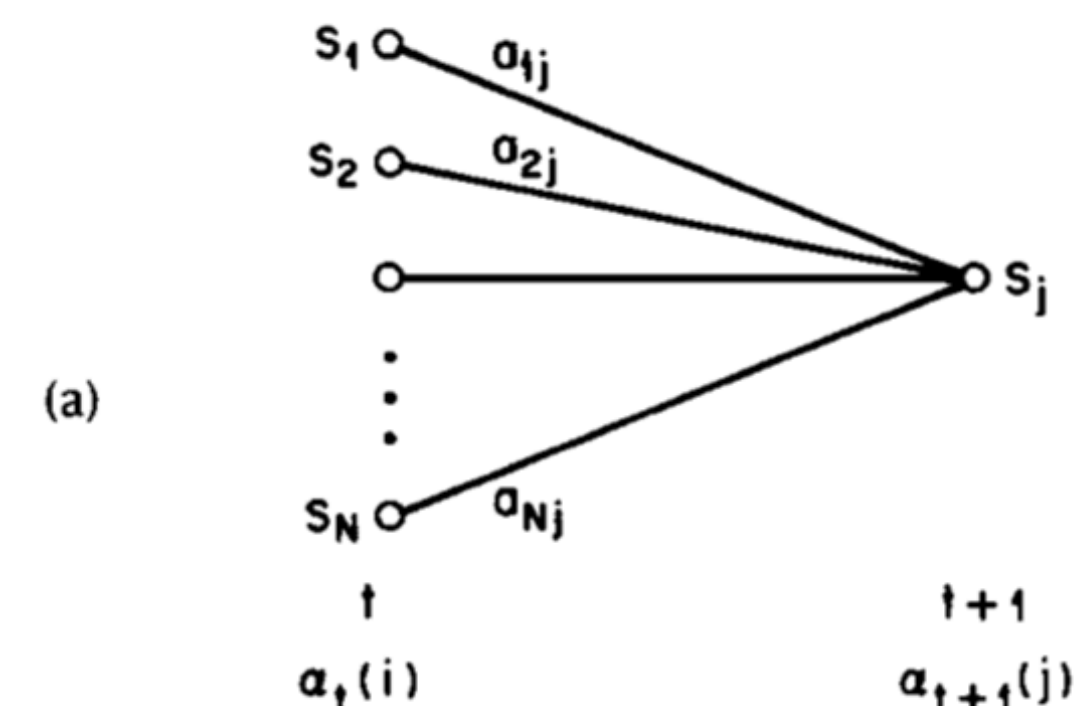
$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \quad (19)$$

2) Induction:

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N. \quad (20)$$

3) Termination:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (21)$$





# Backward variables

In a similar manner,<sup>7</sup> we can consider a backward variable  $\beta_t(i)$  defined as

$$\beta_t(i) = P(O_{t+1} O_{t+2} \cdots O_T | q_t = S_i, \lambda) \quad (23)$$

i.e., the probability of the partial observation sequence from  $t + 1$  to the end, given state  $S_i$  at time  $t$  and the model  $\lambda$ .

Again we can solve for  $\beta_t(i)$  inductively, as follows:

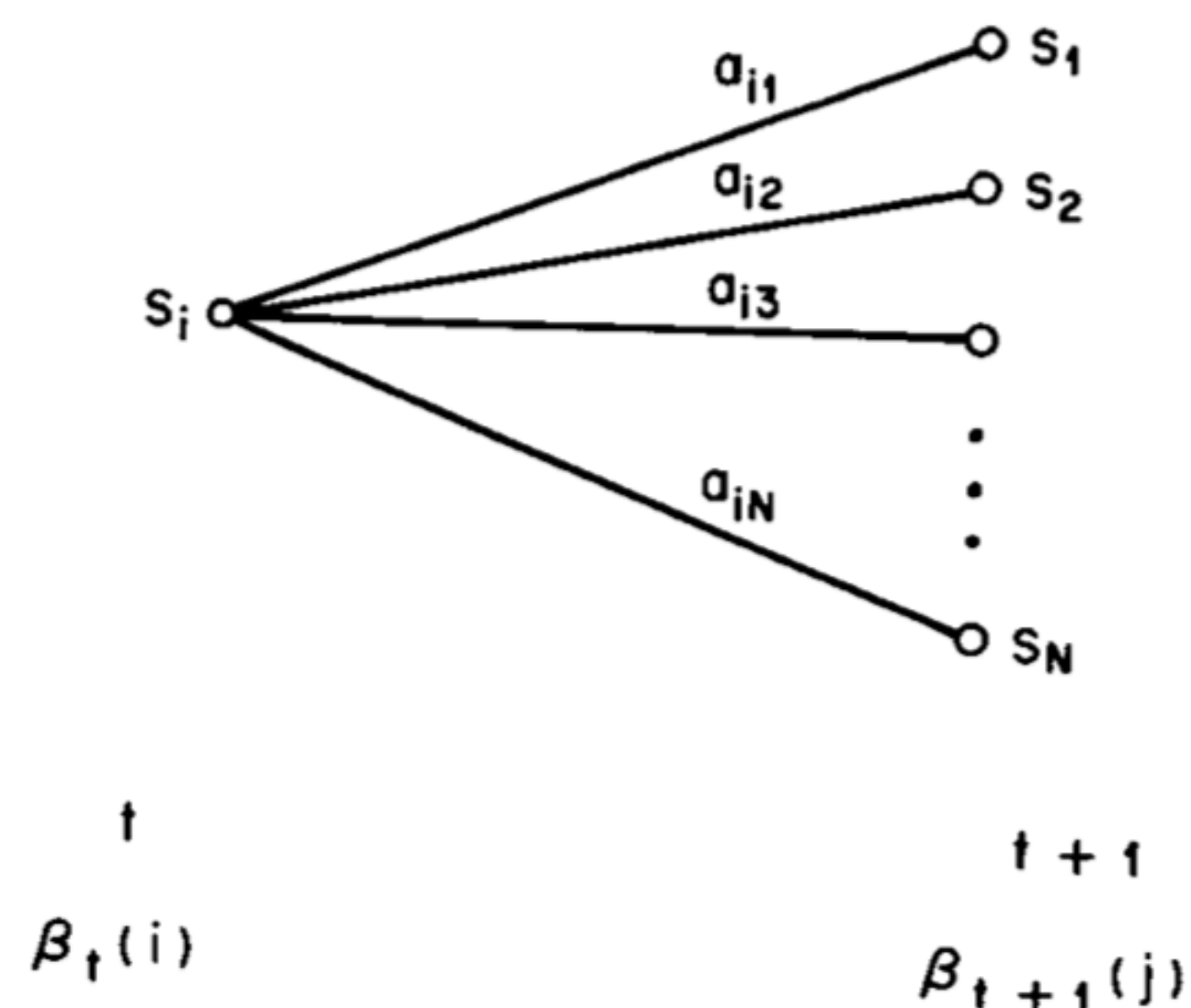
1) Initialization:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N. \quad (24)$$

2) Induction:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j),$$

$$t = T - 1, T - 2, \cdots, 1, 1 \leq i \leq N. \quad (25)$$



# What can we learn from the forward and backward variables?

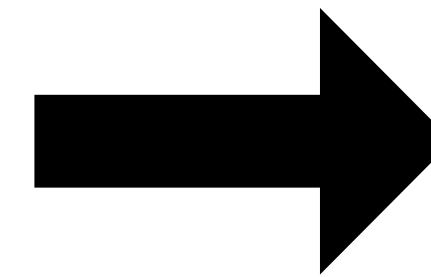
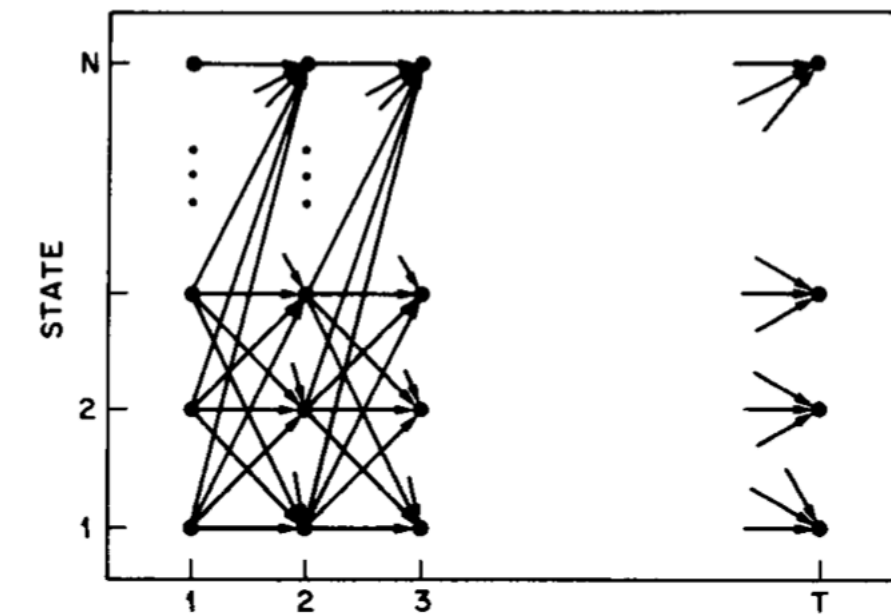
First

We wish to calculate the probability of the observation sequence,  $O = O_1 O_2 \cdots O_T$ , given the model  $\lambda$ , i.e.,  $P(O|\lambda)$ .

The probability of  $O$  (given the model) is obtained by summing this joint probability over all possible state sequences  $q$  giving

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q, \lambda) P(Q|\lambda) \quad (16)$$

$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T). \quad (17)$$



$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i).$$

$2T * N^T$

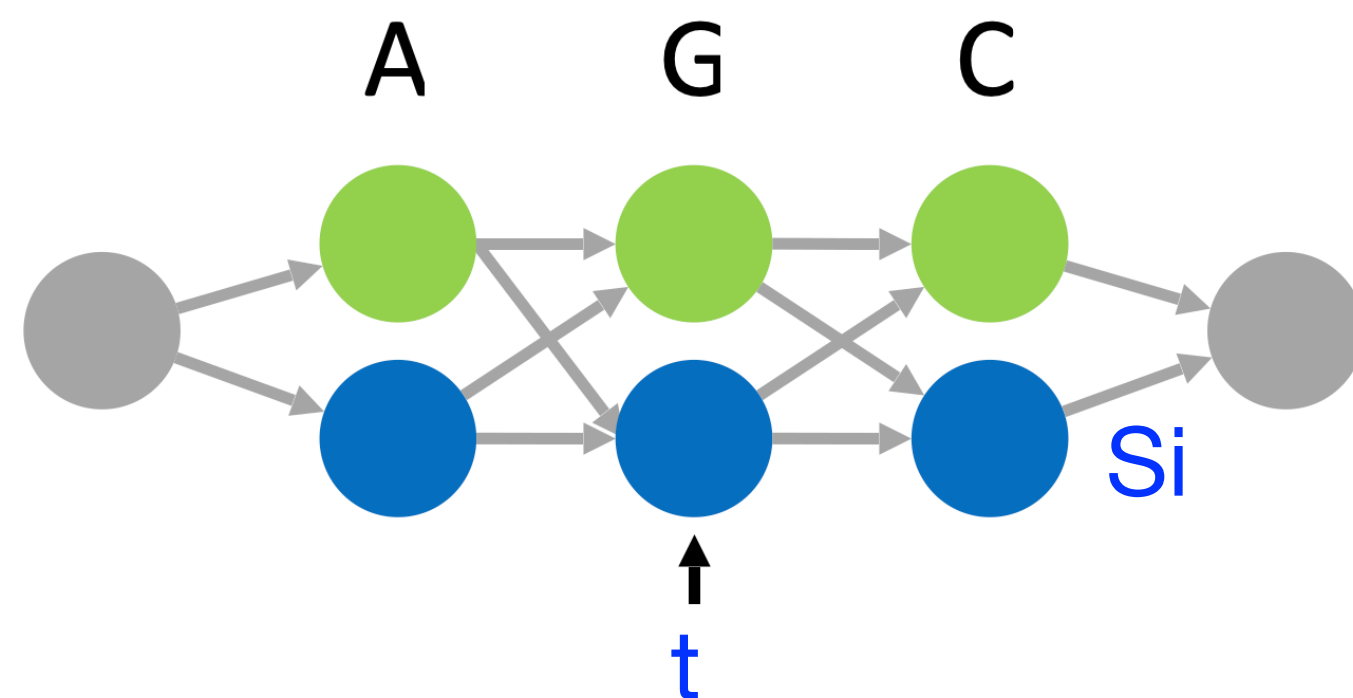
$N^2 * T$

# What can we learn from the forward and backward variables?

## Second

i.e., the probability of being in state  $S_i$  at time  $t$ , given the observation sequence  $O$ , and the model  $\lambda$ . Equation (26) can

$$\gamma_t(i) = P(q_t = S_i | O, \lambda)$$



$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)}$$

$$\sum_{i=1}^N \gamma_t(i) = 1.$$

Using  $\gamma_t(i)$ , we can solve for the individually most likely state  $q_t$  at time  $t$ , as

$$q_t = \operatorname{argmax}_{1 \leq i \leq N} [\gamma_t(i)], \quad 1 \leq t \leq T. \quad (29)$$

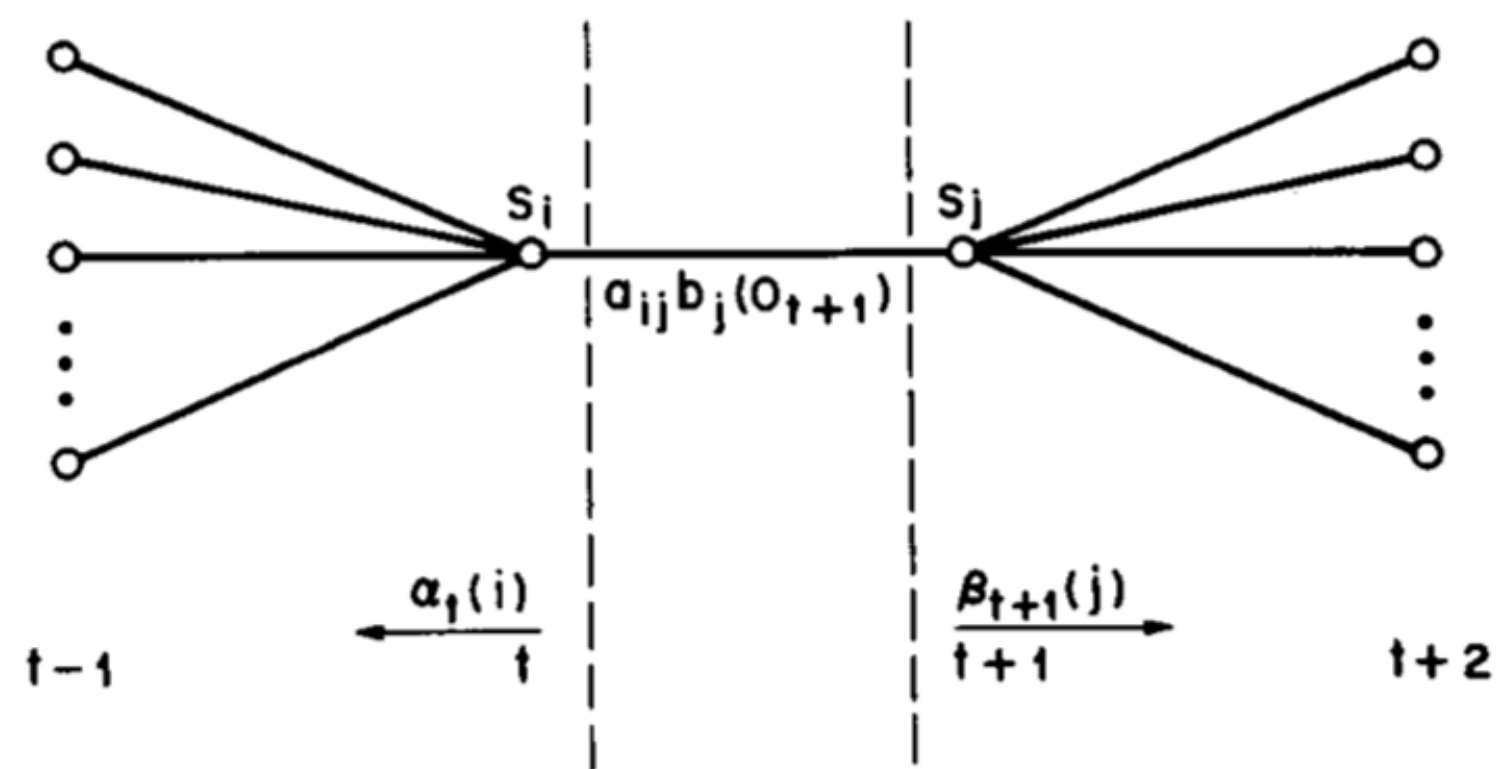
# What can we learn from the forward and backward variables?

Third

first define  $\xi_t(i, j)$ , the probability of being in state  $S_i$  at time  $t$ , and state  $S_j$  at time  $t + 1$ , given the model and the observation sequence, i.e.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda). \quad (36)$$

$$\begin{aligned} \xi_t(i, j) &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned}$$



$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j).$$



# HW8 - Baum-Welch Algorithm

Baum–Welch is an expectation-maximization algorithm that uses the forward–backward algorithm.

1. Use the **forward algorithm** to calculate the forward probabilities for the HMM.
2. Use the **backward algorithm** to calculate the backward probabilities for the HMM.
3. Re-estimate transition, emission, and initial probabilities by calculating the expected number of each edge type
4. Calculate the new log likelihood of the model (the likelihood of our observations given our re-tuned model)
5. Repeat until the change in log likelihood is smaller than a given threshold or when a maximum number of iterations is passed.



# Reestimation for parameters

$\bar{\pi}_i$  = expected frequency (number of times) in state  $S_i$  at time  $(t = 1) = \gamma_1(i)$

$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } S_i \text{ to state } S_j}{\text{expected number of transitions from state } S_i}$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$\bar{b}_j(k) = \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j}$

$$= \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \text{ s.t. } O_t = v_k$$