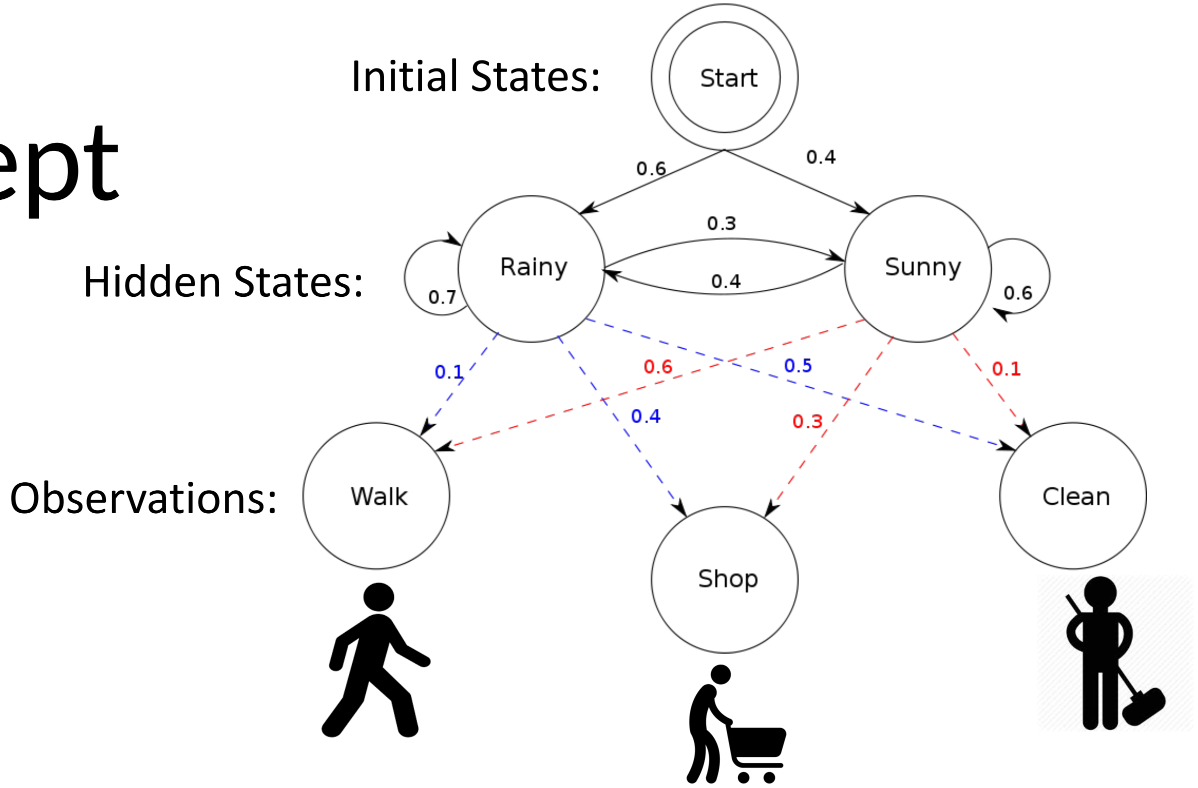
Genome 540 Class 13

Chengxiang Qiu

Agenda

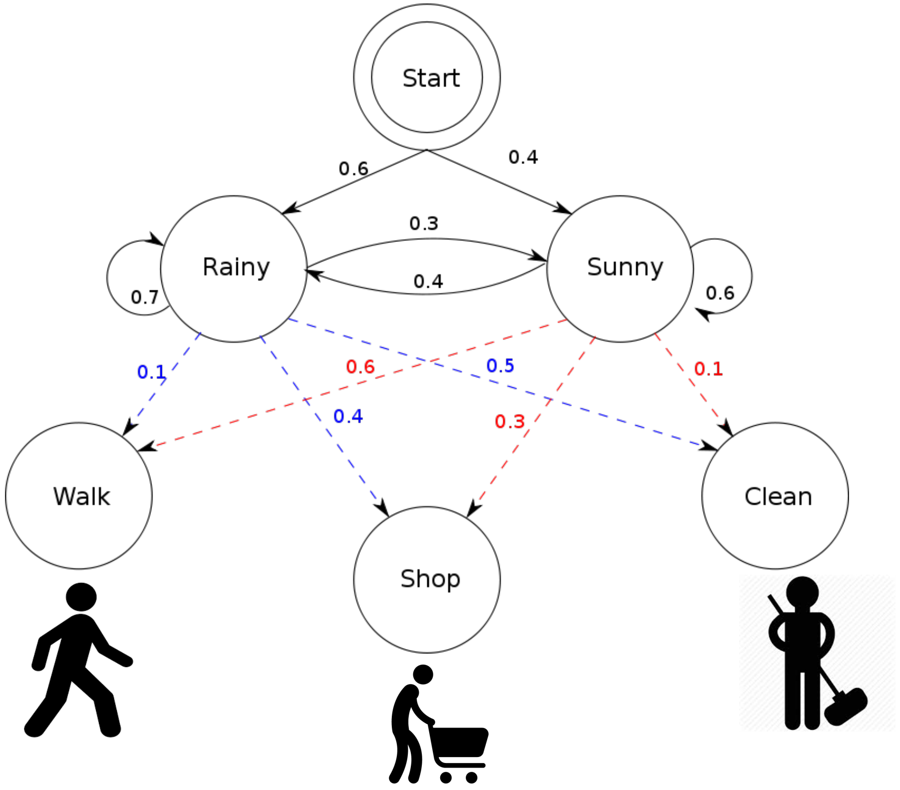
- HW5 questions?
- HMMs: general concept
  - General intuition for how it works
  - Understanding it in a genomic context

HMMs: General Concept



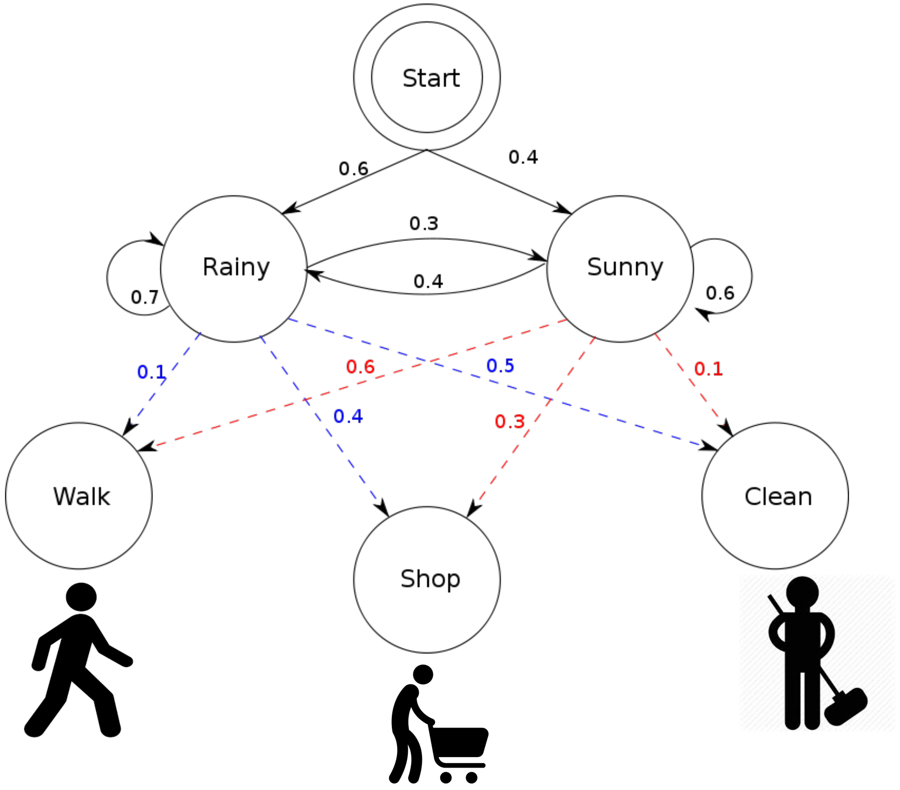
Can we predict what the weather was based on observing a person's behavior?

HMMs: General Concept

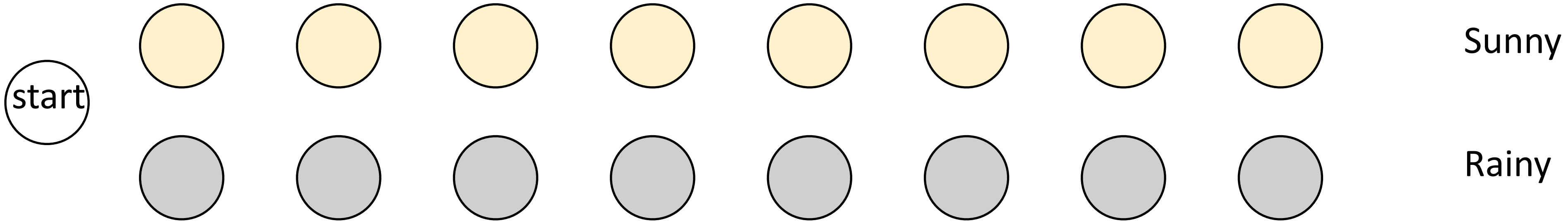


Observation List: Day 1: Walk Day 2: Walk Day 3: Walk Day 4: Walk Day 5: Walk Day 6: Walk Day 7: Walk Day 8: Walk

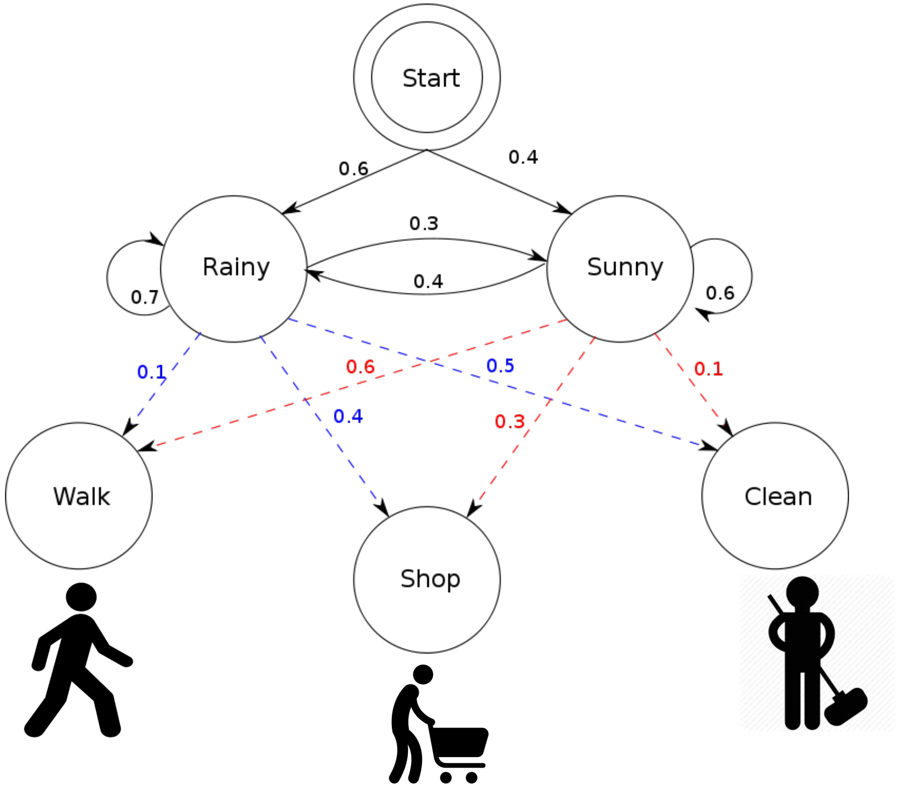
HMMs: General Concept



Observation List: Day 1: Walk Day 2: Walk Day 3: Walk Day 4: Walk Day 5: Walk Day 6: Walk Day 7: Walk Day 8: Walk

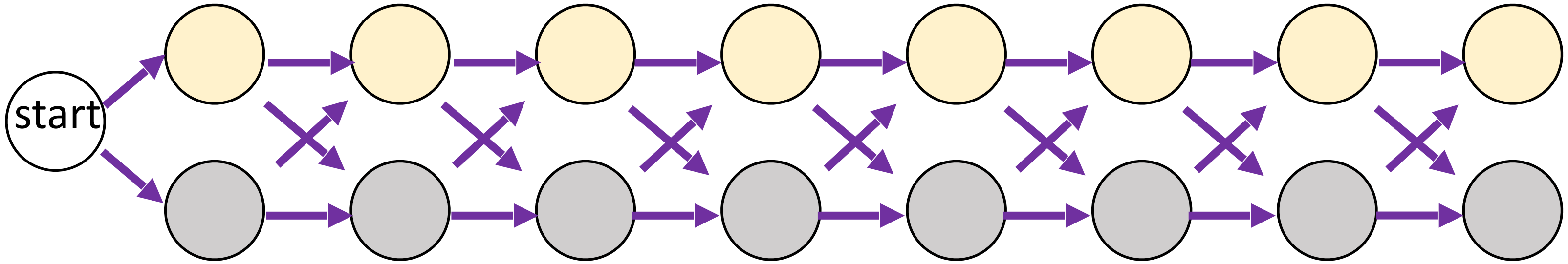


HMMs: General Concept



Observation List: Day 1: Walk, Day 2: Walk, Day 3: Walk, Day 4: Walk, Day 5: Walk, Day 6: Walk, Day 7: Walk, Day 8: Walk

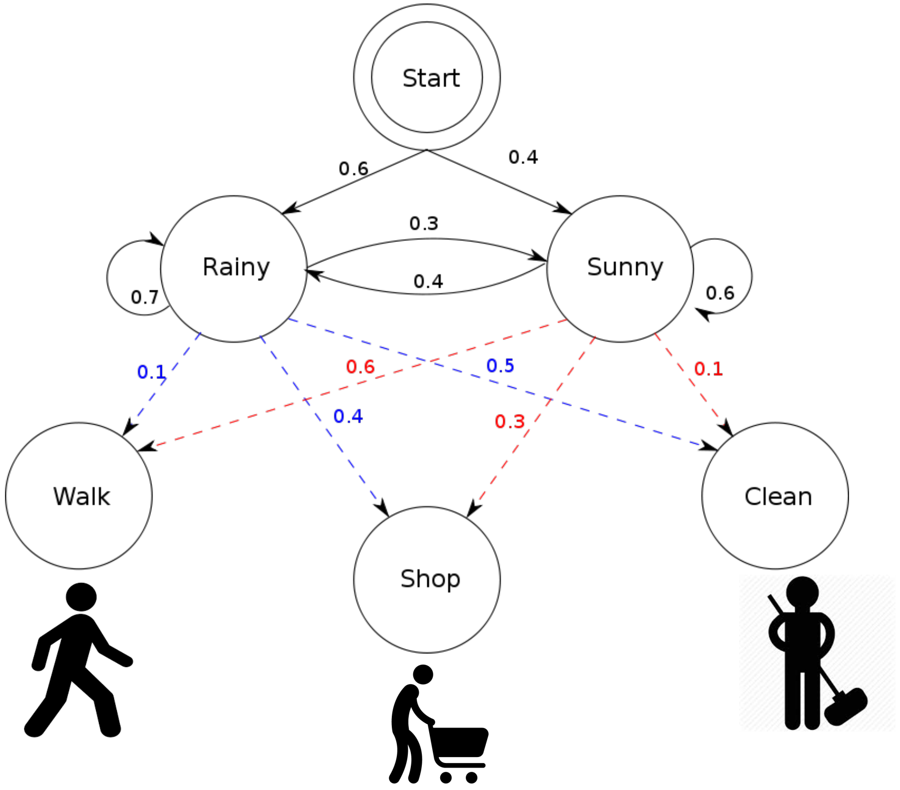
All possible paths



Sunny

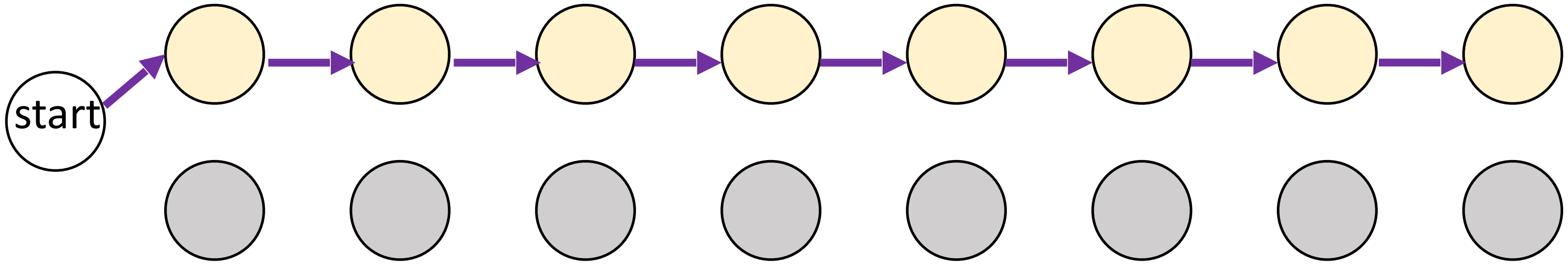
Rainy

HMMs: General Concept



Observation List: Day 1: Walk, Day 2: Walk, Day 3: Walk, Day 4: Walk, Day 5: Walk, Day 6: Walk, Day 7: Walk, Day 8: Walk

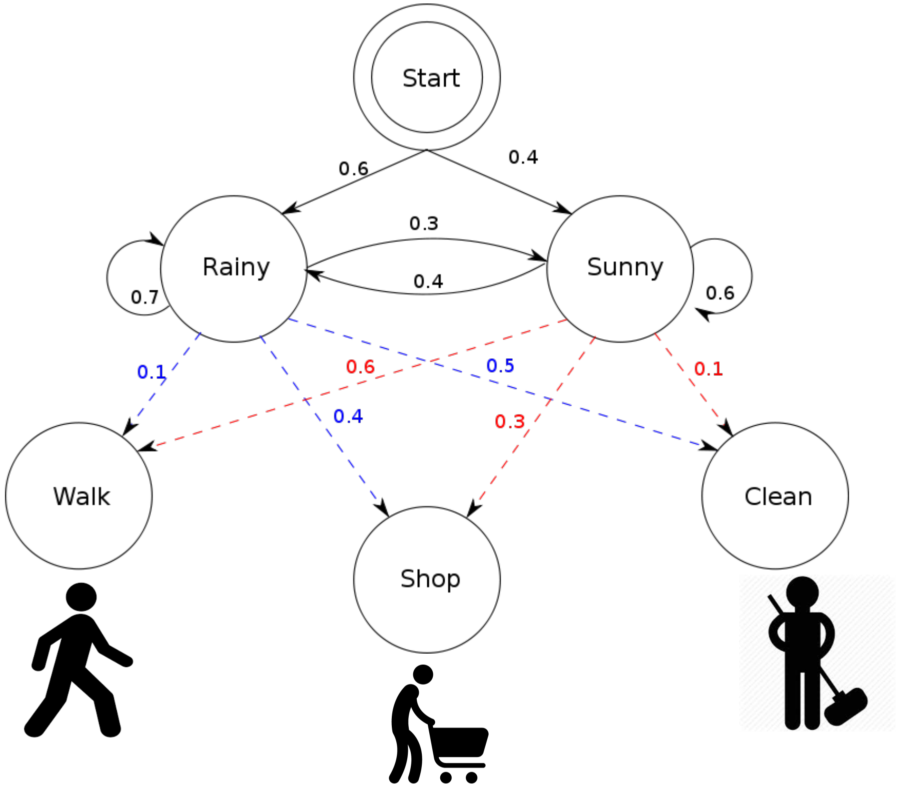
Highest Weight Path (i.e. most probable path)



Sunny

Rainy

HMMs: General Concept



What is the most probable path now?

Observation List:

Day 1: Clean

Day 2: Clean

Day 3: Clean

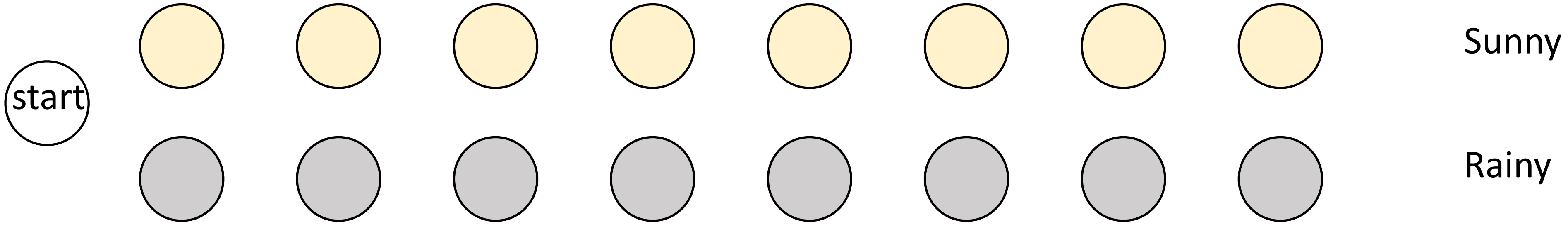
Day 4: Clean

Day 5: Clean

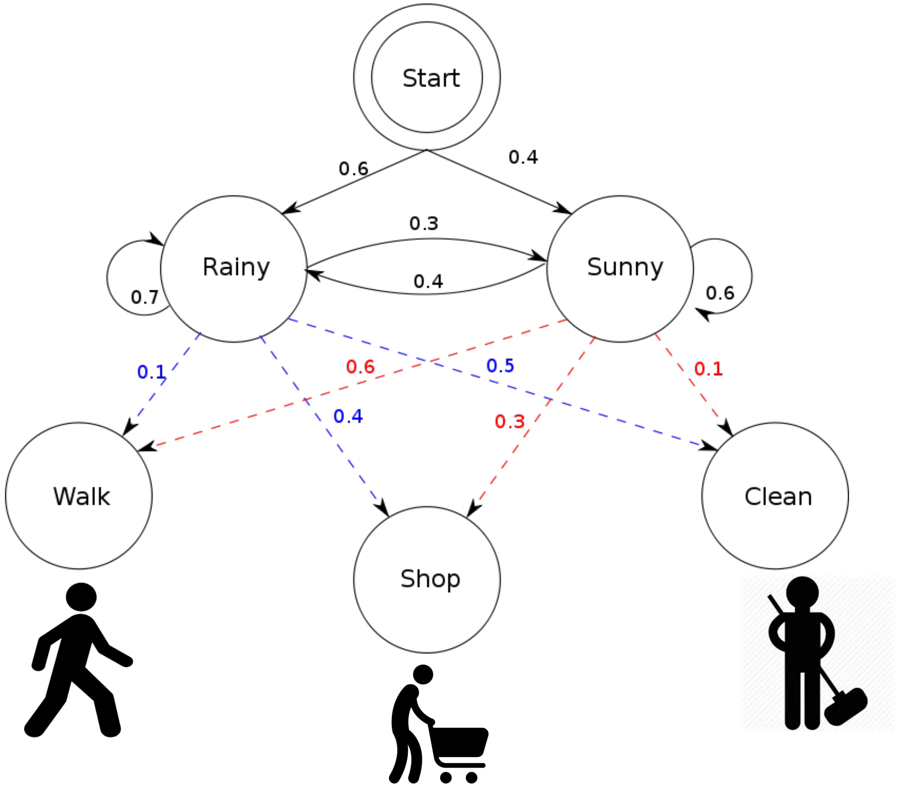
Day 6: Clean

Day 7: Clean

Day 8: Clean



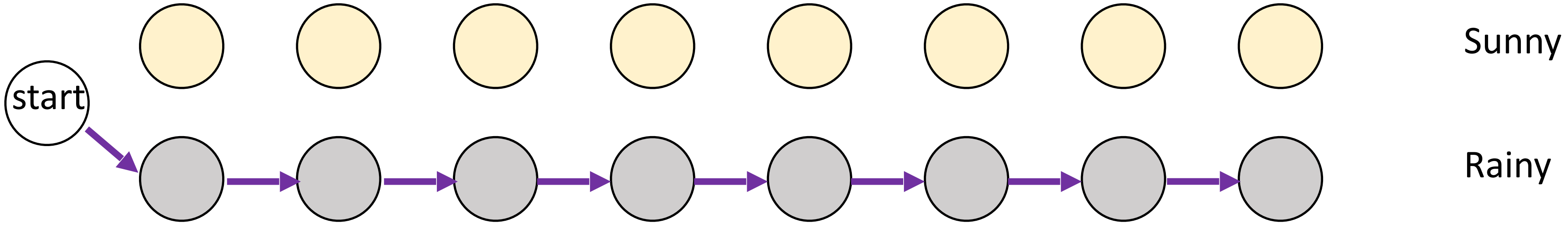
HMMs: General Concept



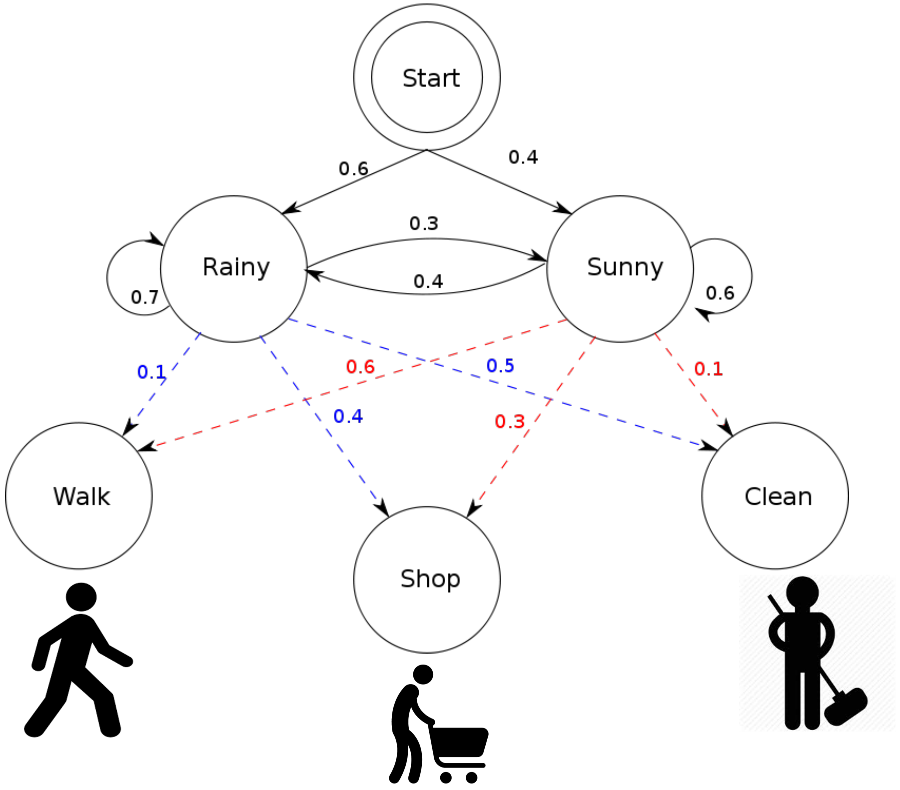
What is the most probable path now?

Observation List:

Day 1: Clean
Day 2: Clean
Day 3: Clean
Day 4: Clean
Day 5: Clean
Day 6: Clean
Day 7: Clean
Day 8: Clean

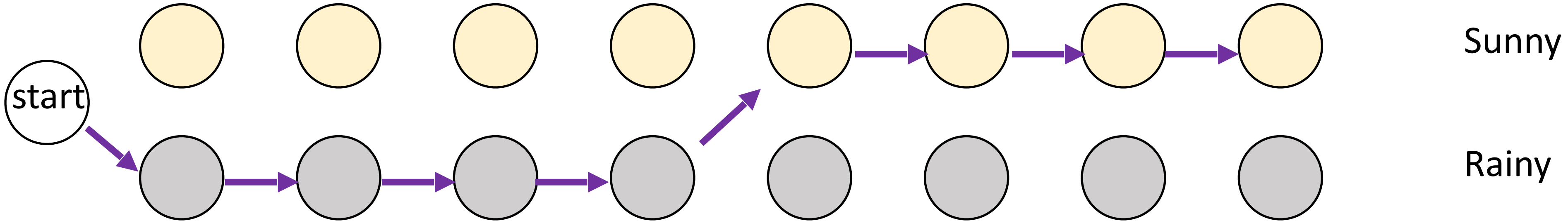


HMMs: General Concept

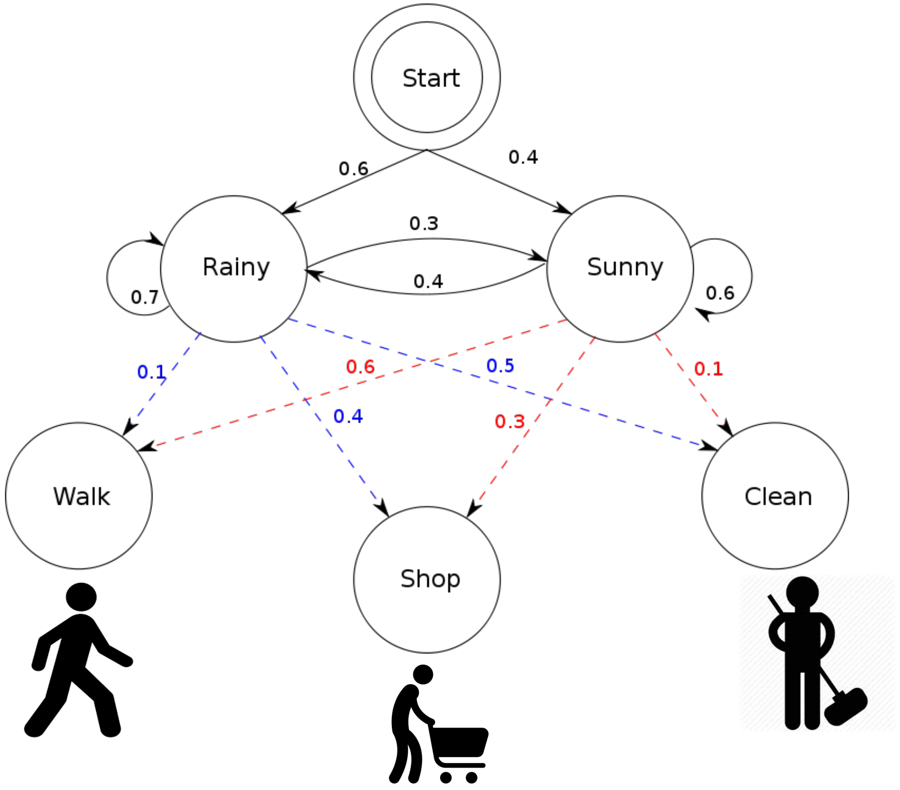


What is the most probable path now?

Observation List: Day 1: Clean, Day 2: Clean, Day 3: Clean, Day 4: Clean, Day 5: Walk, Day 6: Walk, Day 7: Walk, Day 8: Walk

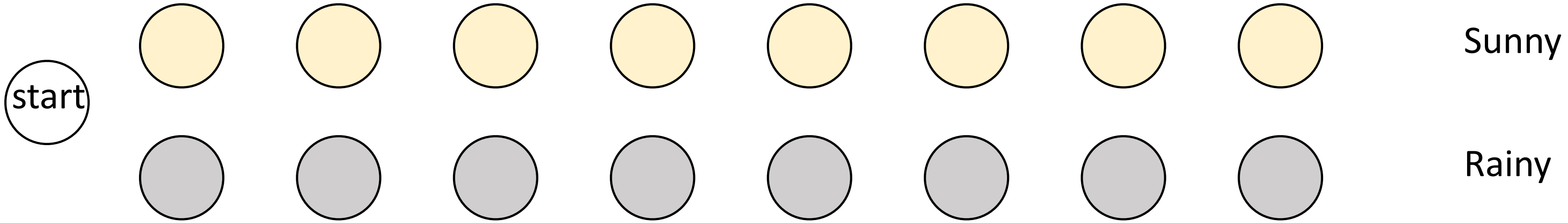


HMMs: General Concept

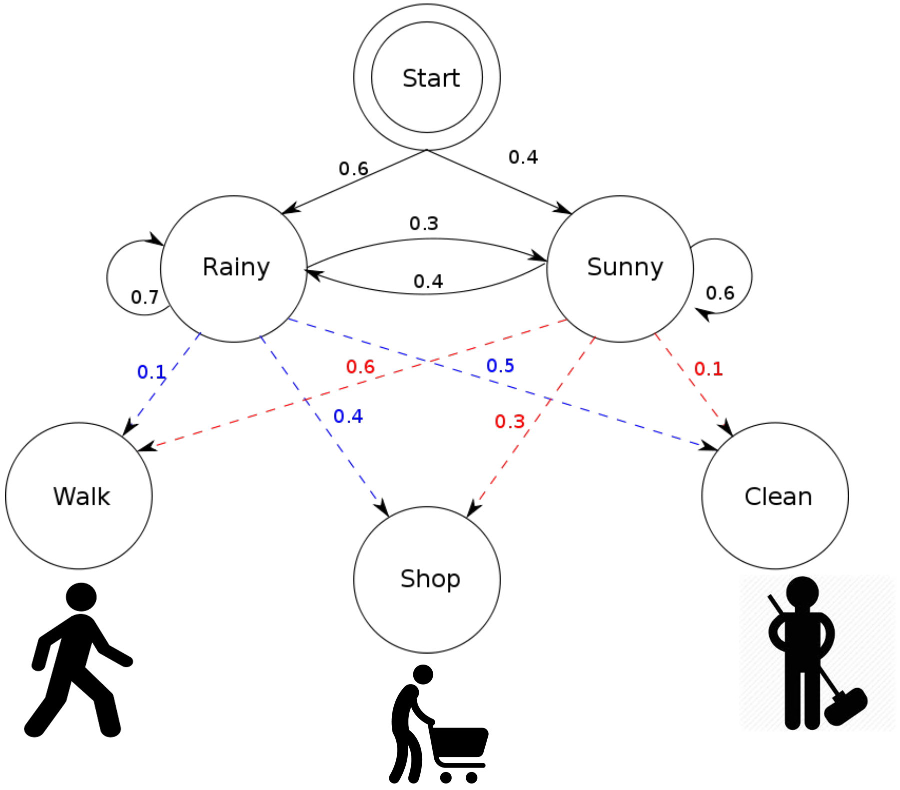


What about more complicated observations?

Observation List: Day 1: Shop, Day 2: Walk, Day 3: Clean, Day 4: Walk, Day 5: Shop, Day 6: Walk, Day 7: Clean, Day 8: Shop



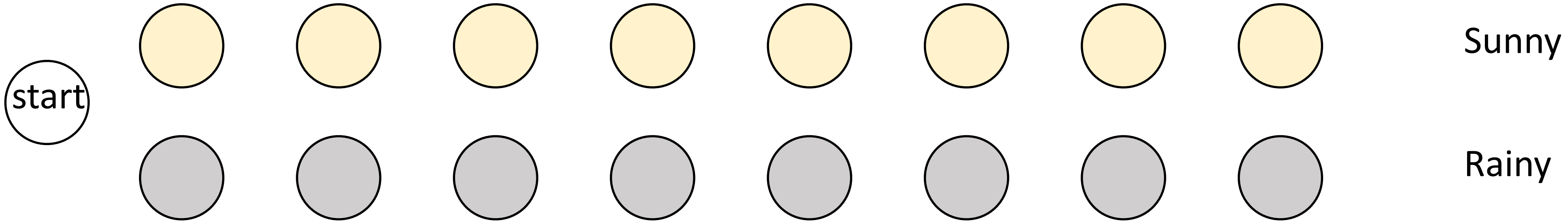
HMMs: General Concept



What about more complicated observations?

We now need an algorithm to crunch the probability of all possible paths.

Observation List: Day 1: Shop, Day 2: Walk, Day 3: Clean, Day 4: Walk, Day 5: Shop, Day 6: Walk, Day 7: Clean, Day 8: Shop



A general definition of HMM

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

N states

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

M symbols

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

Initial state probability:

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

$$\pi_i : \textit{prob of starting at state } s_i$$

State transition probability:

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

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

Output probability:

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

$$b_i(v_k) : \textit{prob of "generating" } v_k \textit{ at } s_i$$

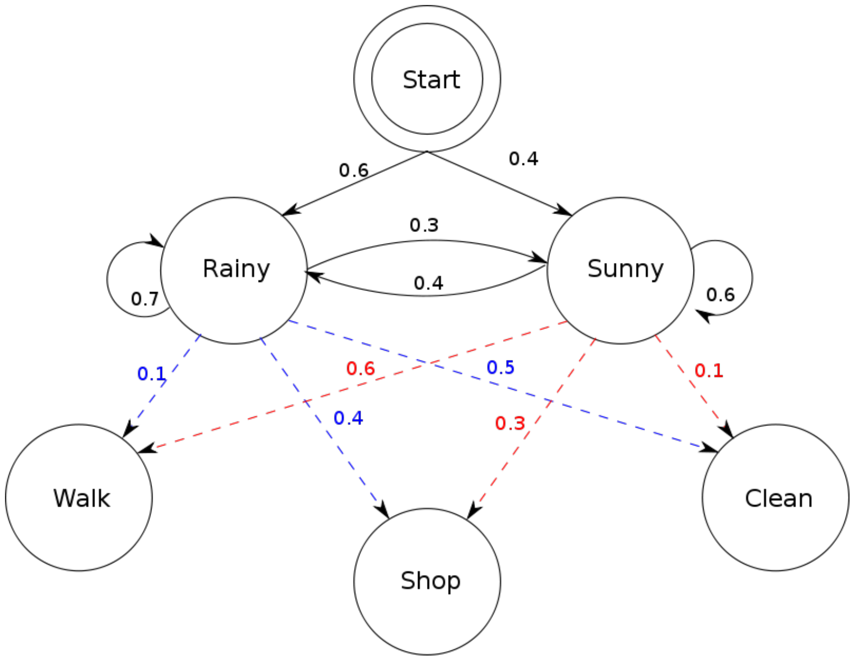
Three Basic Problems in HMMs

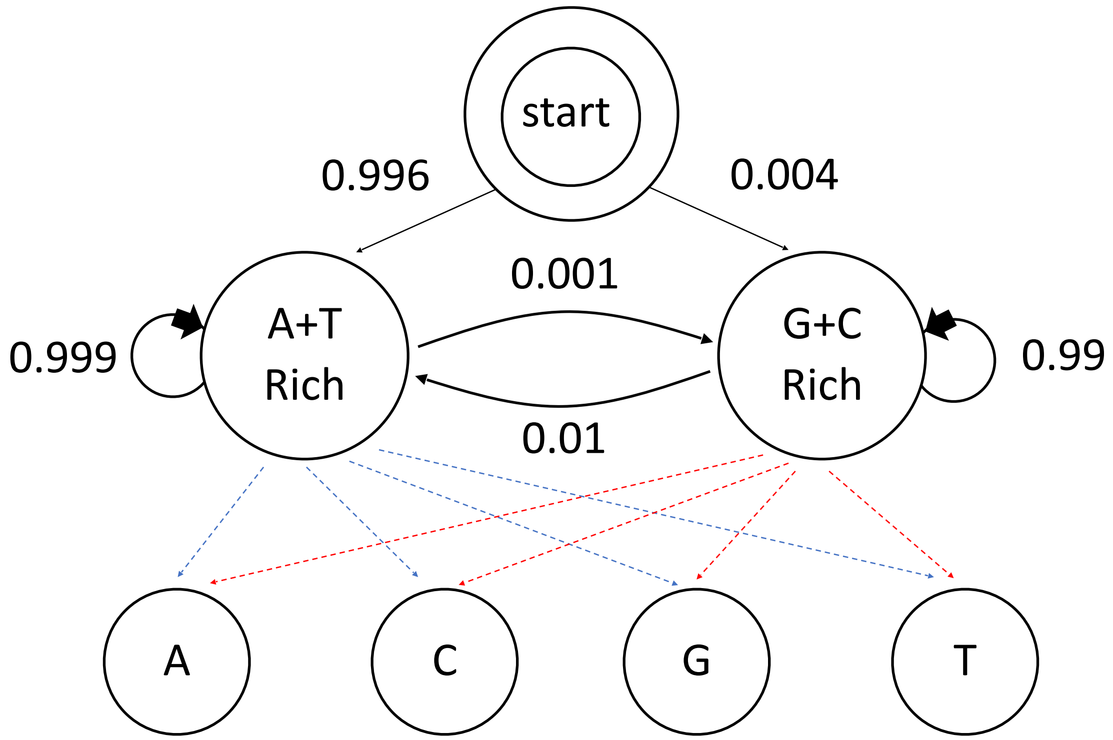
Given a set of observation sequences $O = O_1 O_2 \cdots O_T$ and the HMM parameters $\lambda = (A, B, \pi)$, computing the probability $P(O|\lambda)$

Given a set of observation sequences $O = O_1 O_2 \cdots O_T$ and the HMM parameters $\lambda = (A, B, \pi)$, computing the optimal state sequences

Given a set of observation sequences $O = O_1 O_2 \cdots O_T$ adjusting the HMM parameters $\lambda = (A, B, \pi)$ to maximize the probability $P(O|\lambda)$

Applying this Concept To Genomics/HW

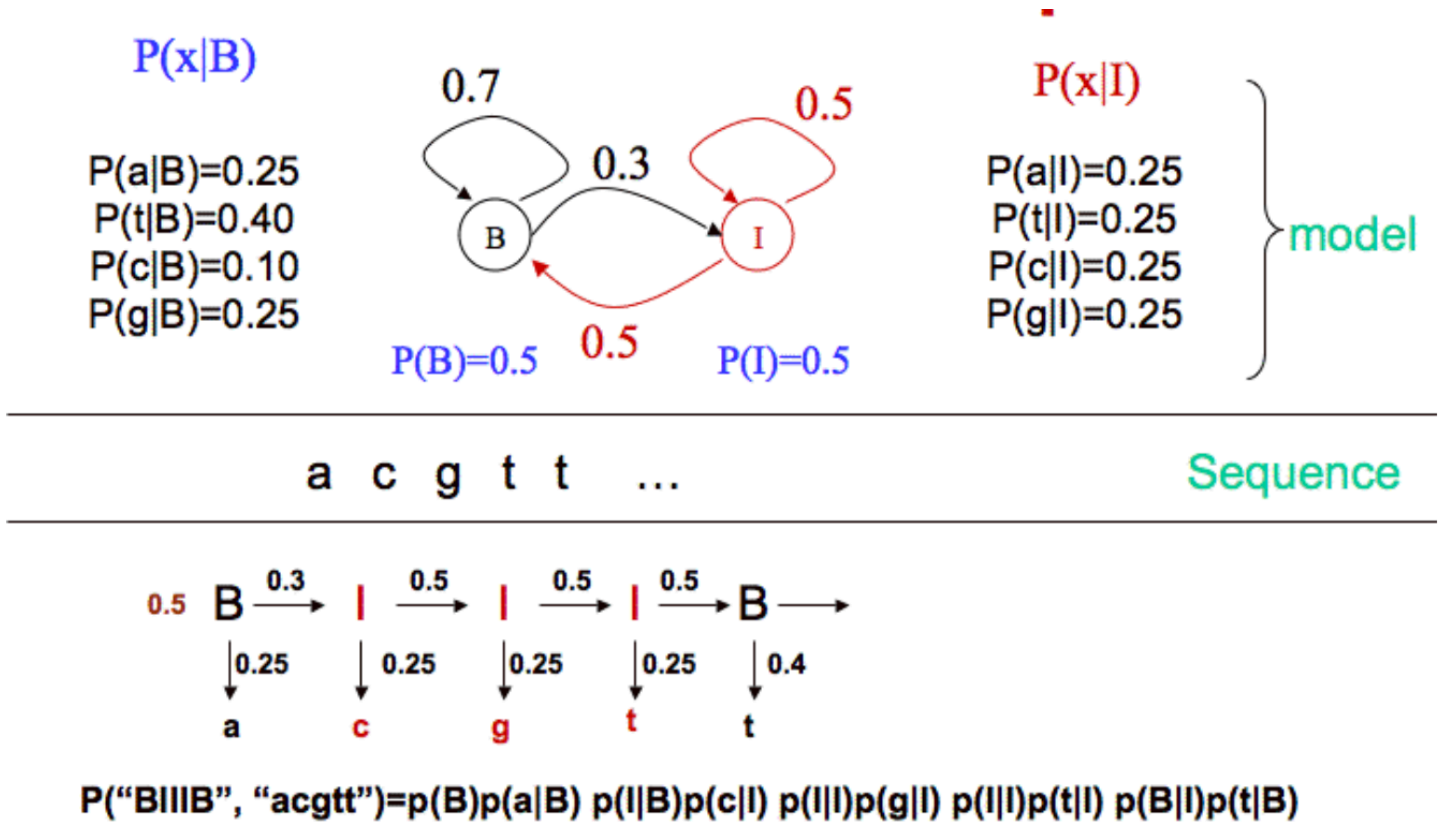




A = 0.291
C = 0.209
G = 0.209
T = 0.291

A = 0.169
C = 0.331
G = 0.331
T = 0.169

How to “generate” a sequence?



What's the most likely path?