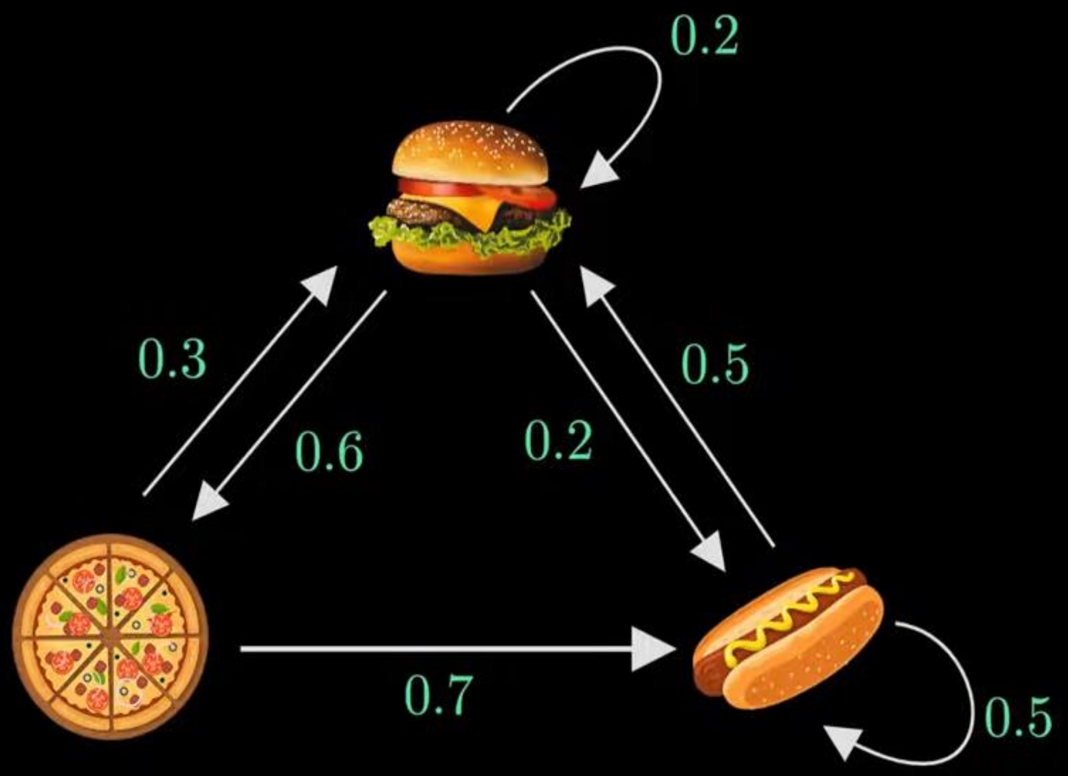
Hidden Markov Model

Zahoor Tanoli (PhD)
CUI Attock

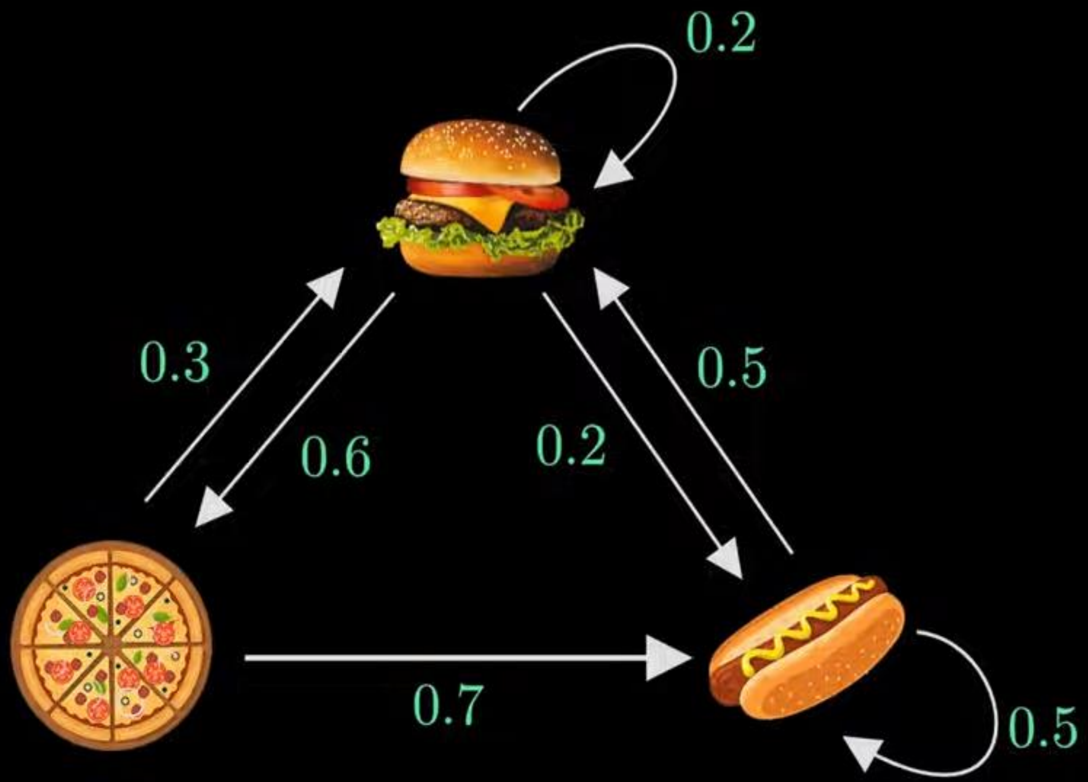
Markov Chain :

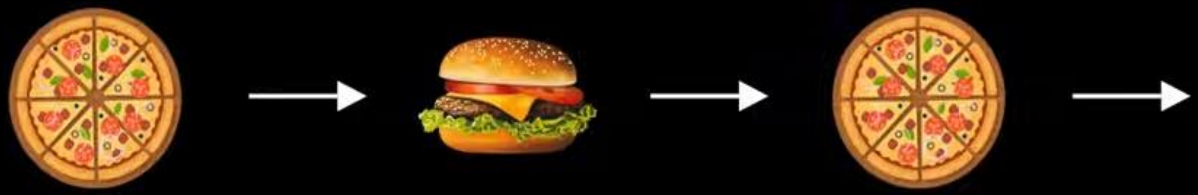
a process with a finite number of states (or outcomes, or events) in which the probability of being in a particular state at step n + 1 depends only on the state occupied at step n.

Prof. Andrei A. Markov (1856-1922) , published his result in 1906.



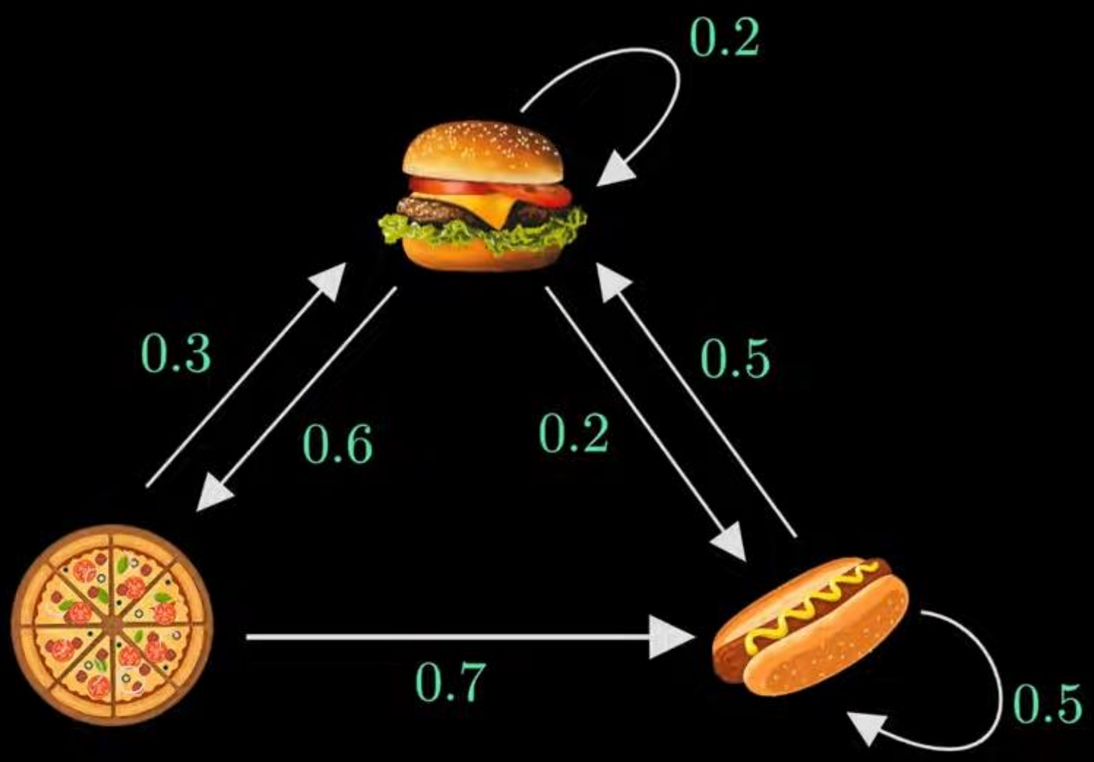
$$P(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$





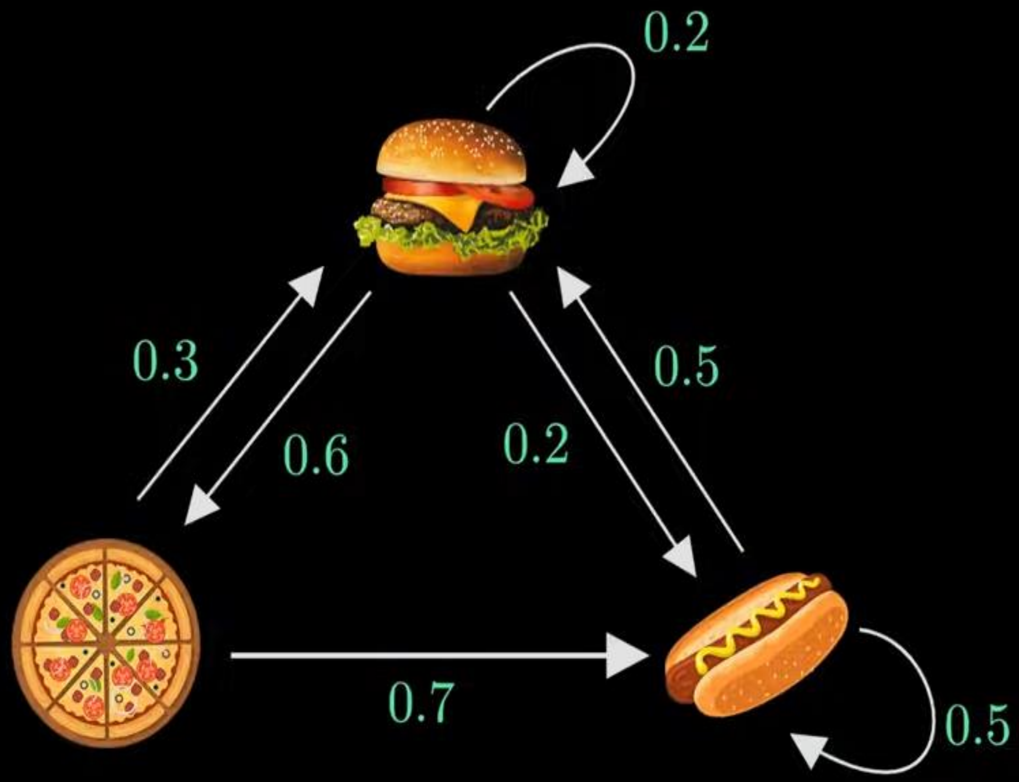
$$P(X_{n+1} = x \mid X_n = x_n)$$

$P(X_4 =$  $\mid X_1 =$  $, X_2 =$  $, X_3 =$  $)$



$$P(X_{n+1} = x \mid X_n = x_n)$$

$P(X_4 =$  $\mid X_3 =$  $) = 0.7$



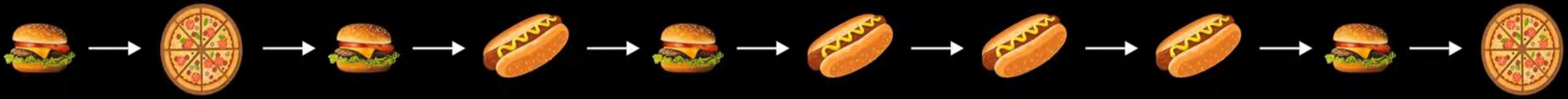
After 10 steps...

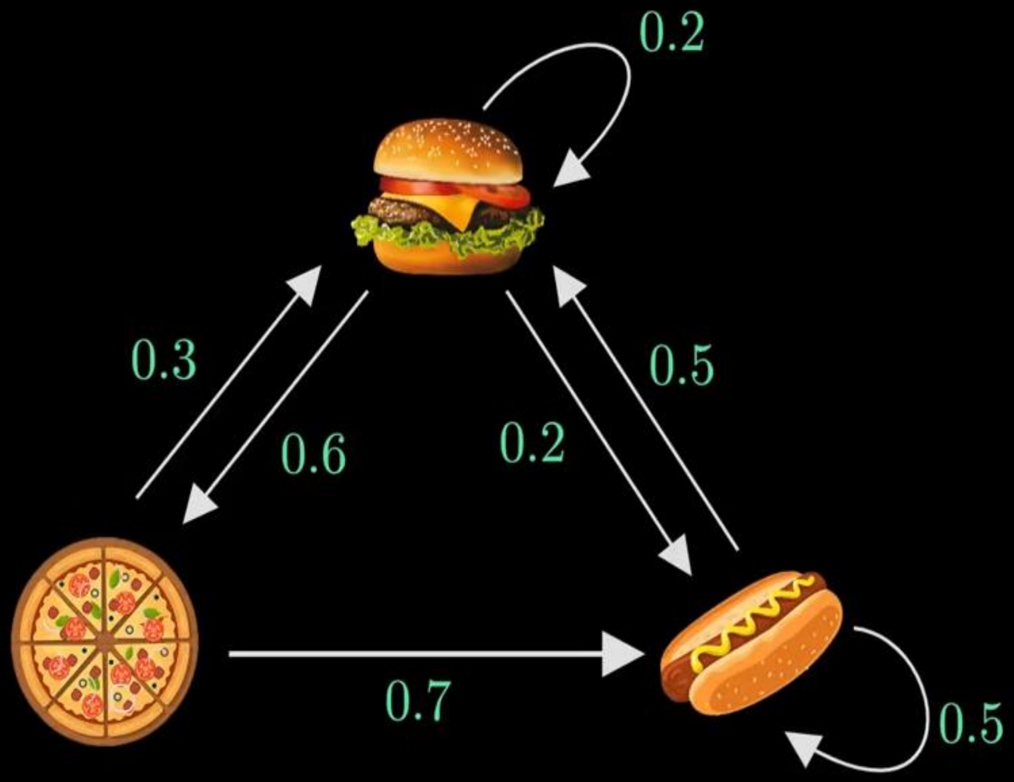
$P($  $)$ $\frac{4}{10}$

$P($  $)$ $\frac{2}{10}$

$P($  $)$ $\frac{4}{10}$

Random Walk

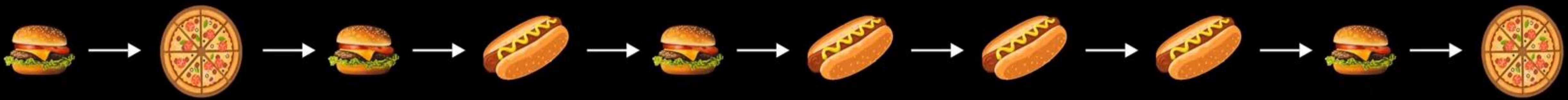


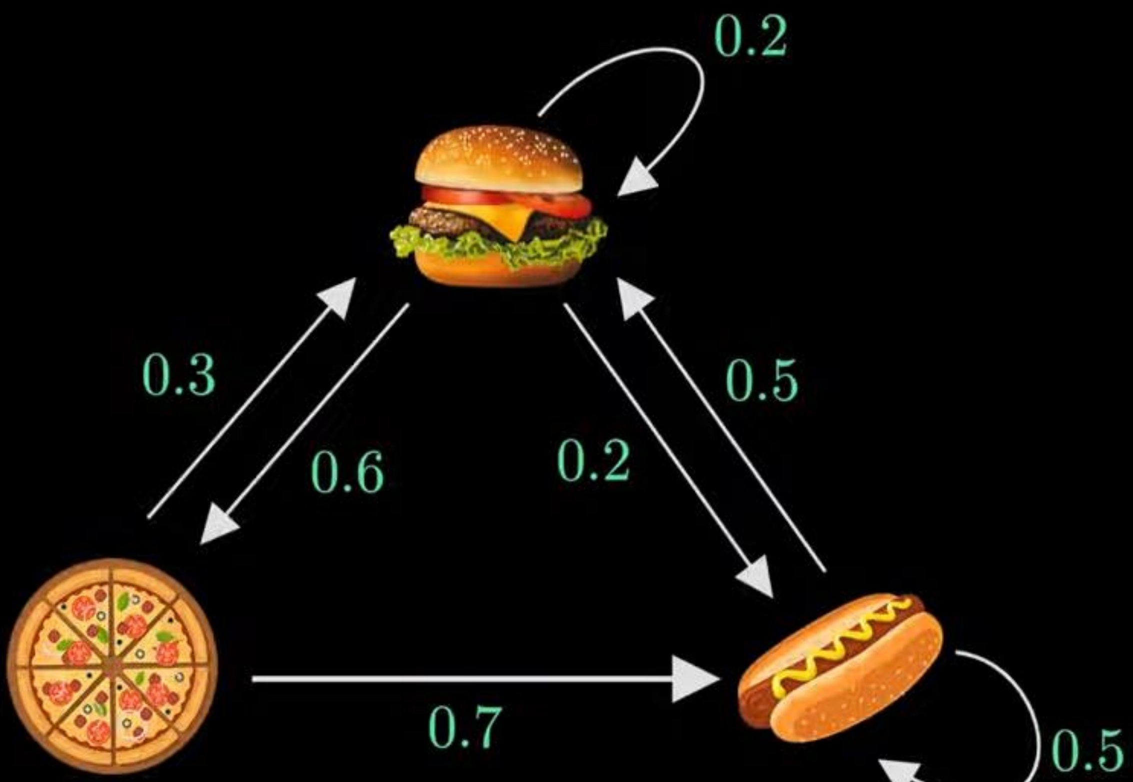


After ∞ steps.. **Stationary/Equilibrium State**

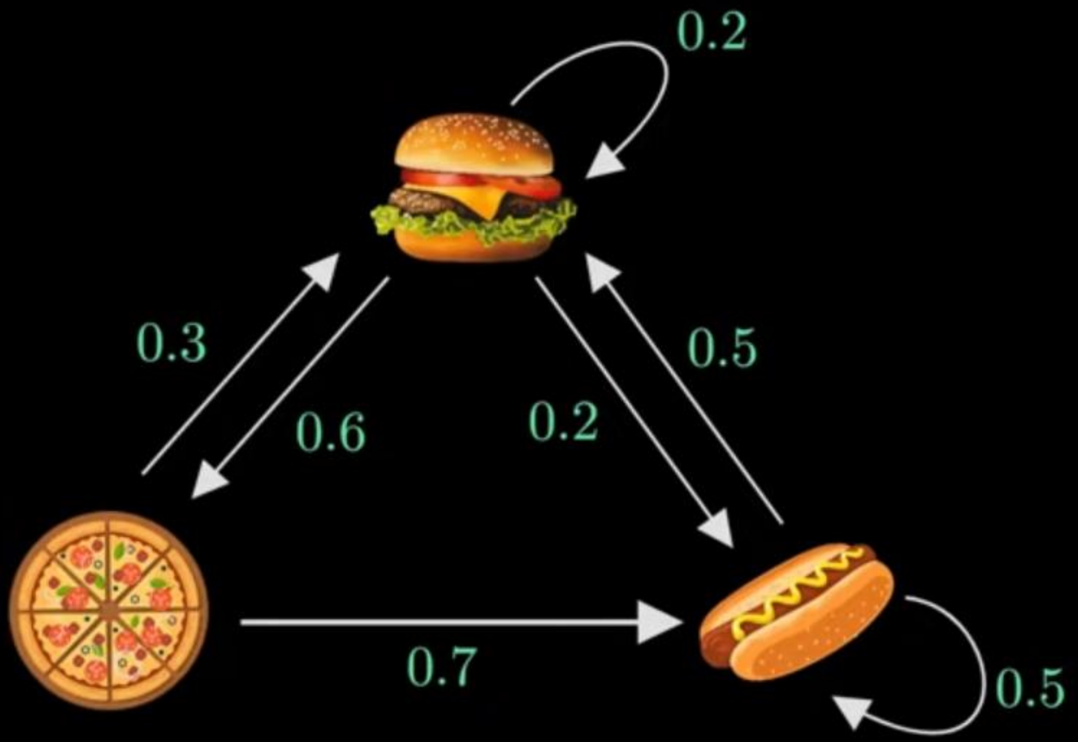
$P($$)$	$P($$)$	$P($$)$
0.35191	0.21245	0.43564

Random Walk





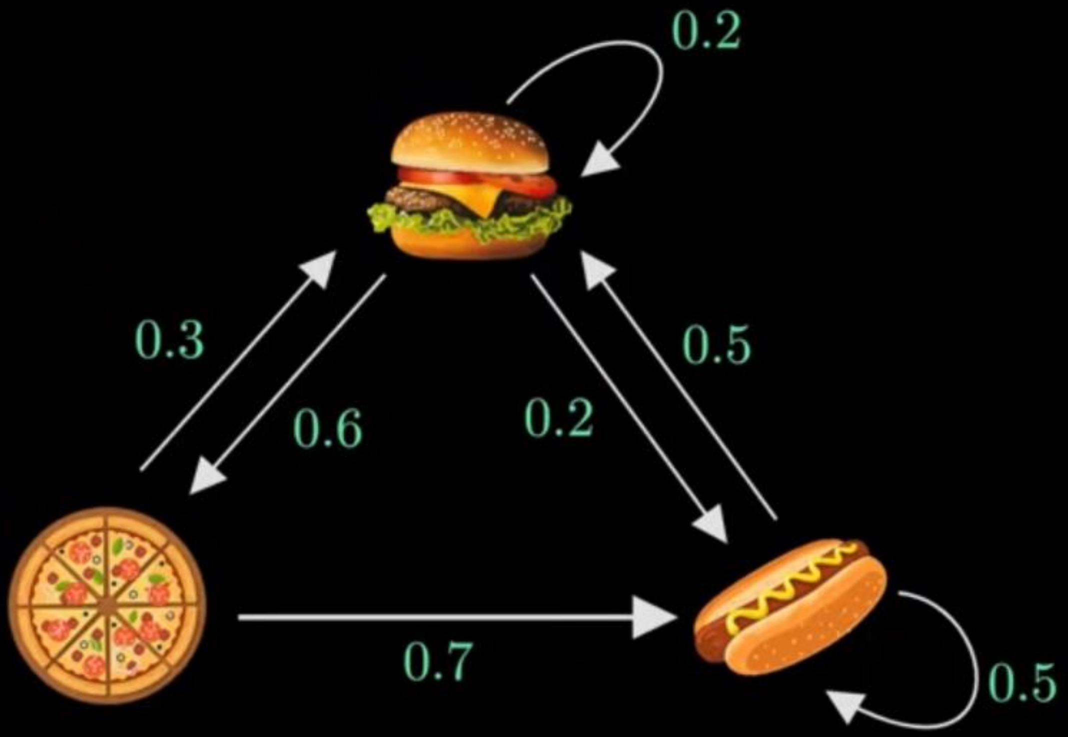
$$\begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$



$$A = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

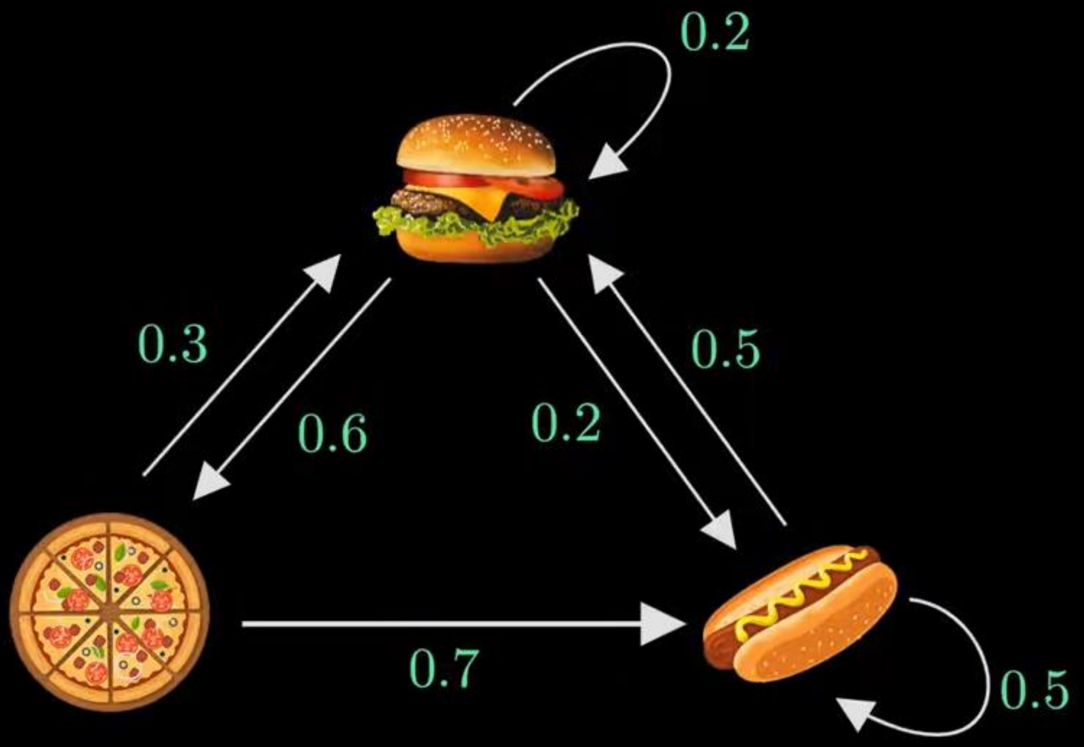
$$\pi_0 A = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.3 & 0 & 0.7 \end{bmatrix}$$



$$A = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$\pi_1 A = \begin{bmatrix} 0.3 & 0 & 0.7 \end{bmatrix} \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.41 & 0.18 & 0.41 \end{bmatrix}$$

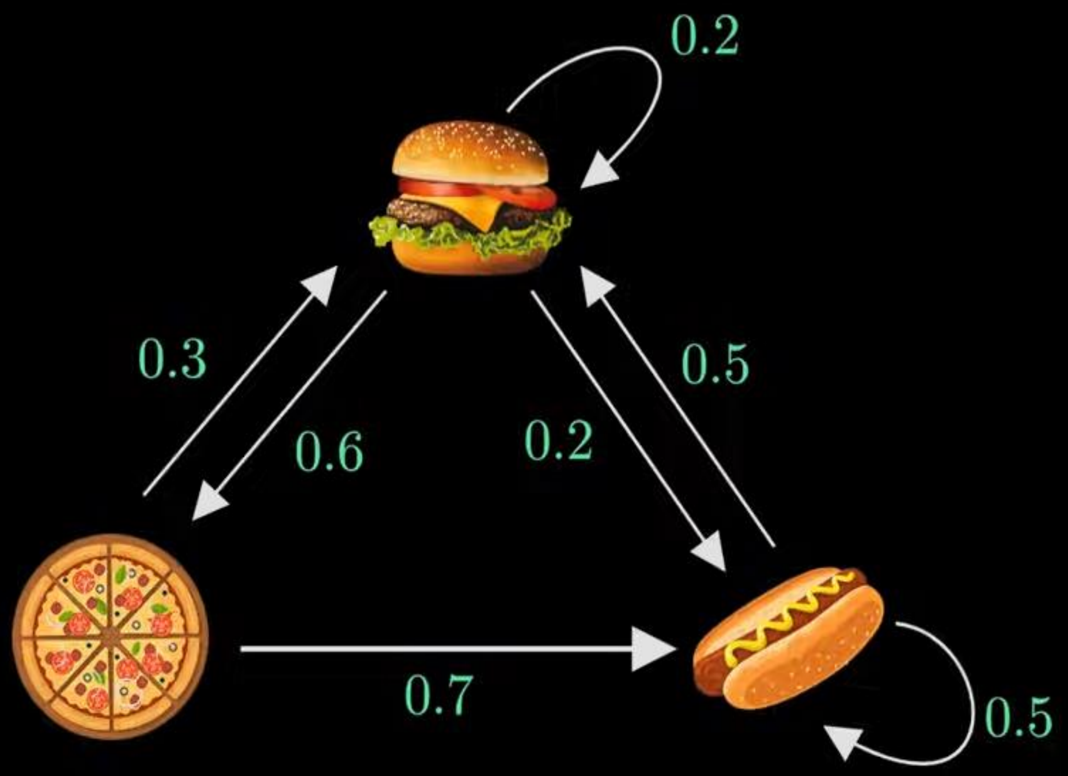


$$A = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

Stationary state

$$\pi_2 A = \begin{bmatrix} 0.41 & 0.18 & 0.41 \end{bmatrix} \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.34 & 0.25 & 0.41 \end{bmatrix}$$

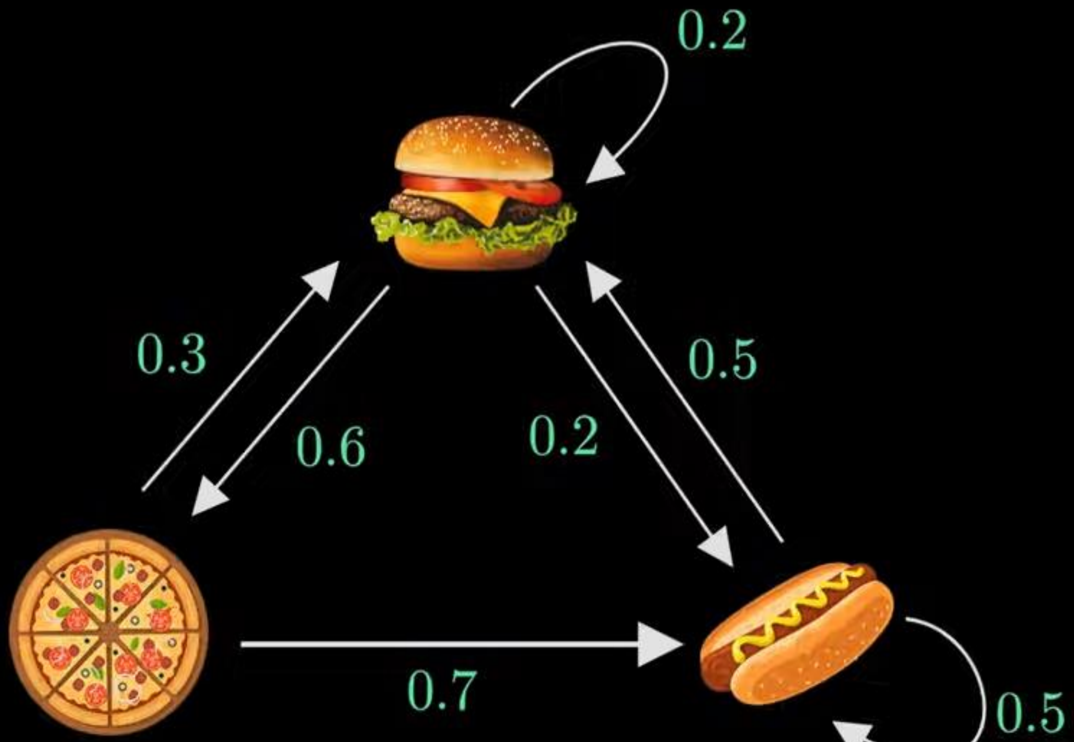


$$A = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$\pi A = \pi$$

$$Av = \lambda v$$

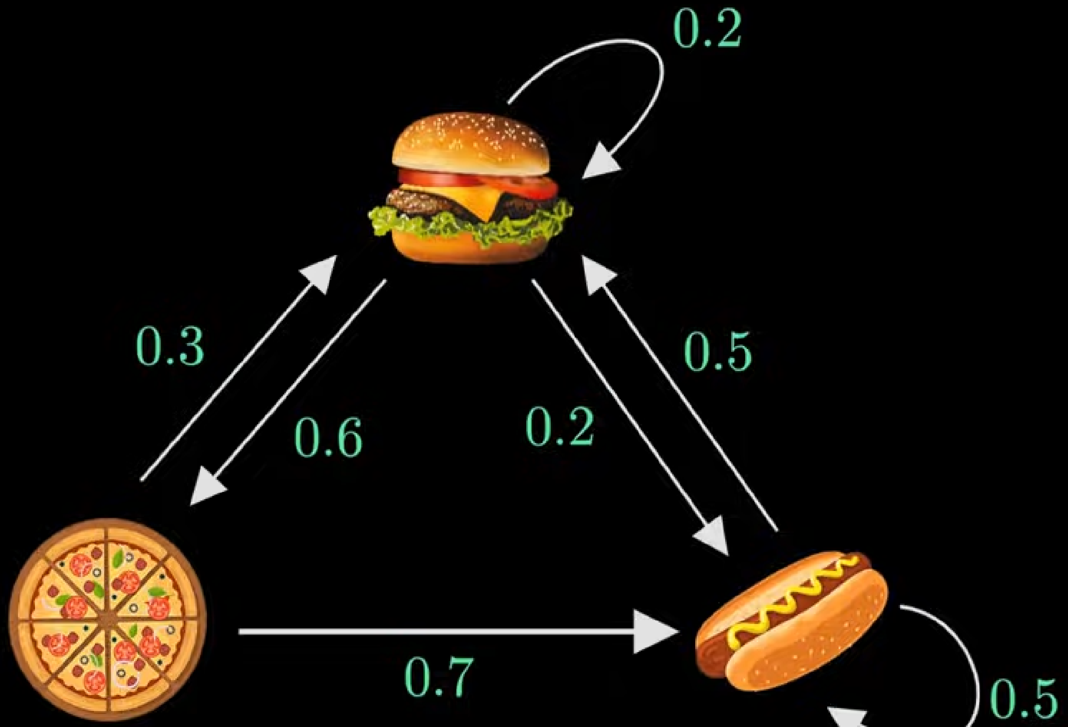


$$A = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$\pi A = \pi$$

$$\pi[1] + \pi[2] + \pi[3] = 1$$

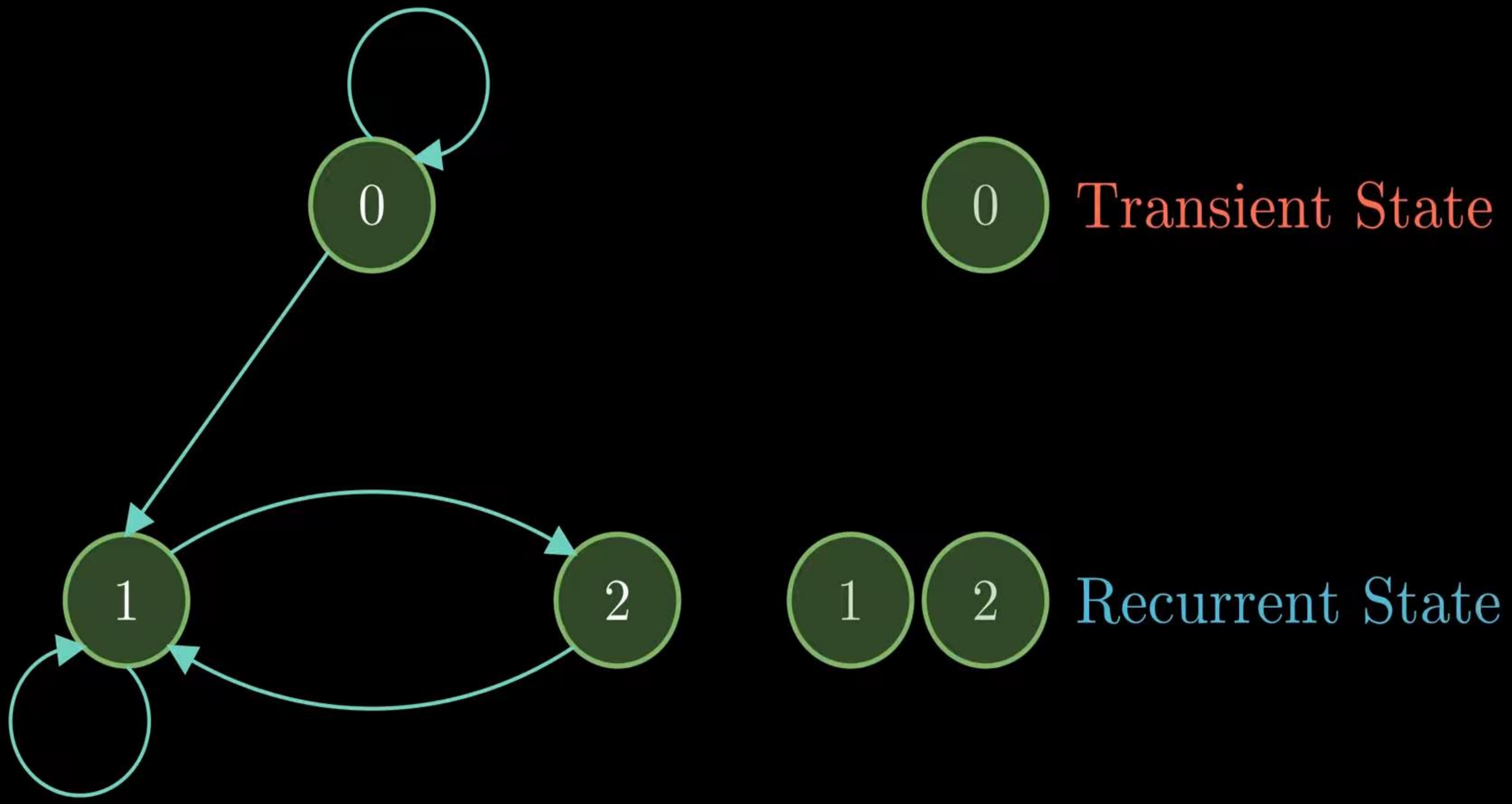


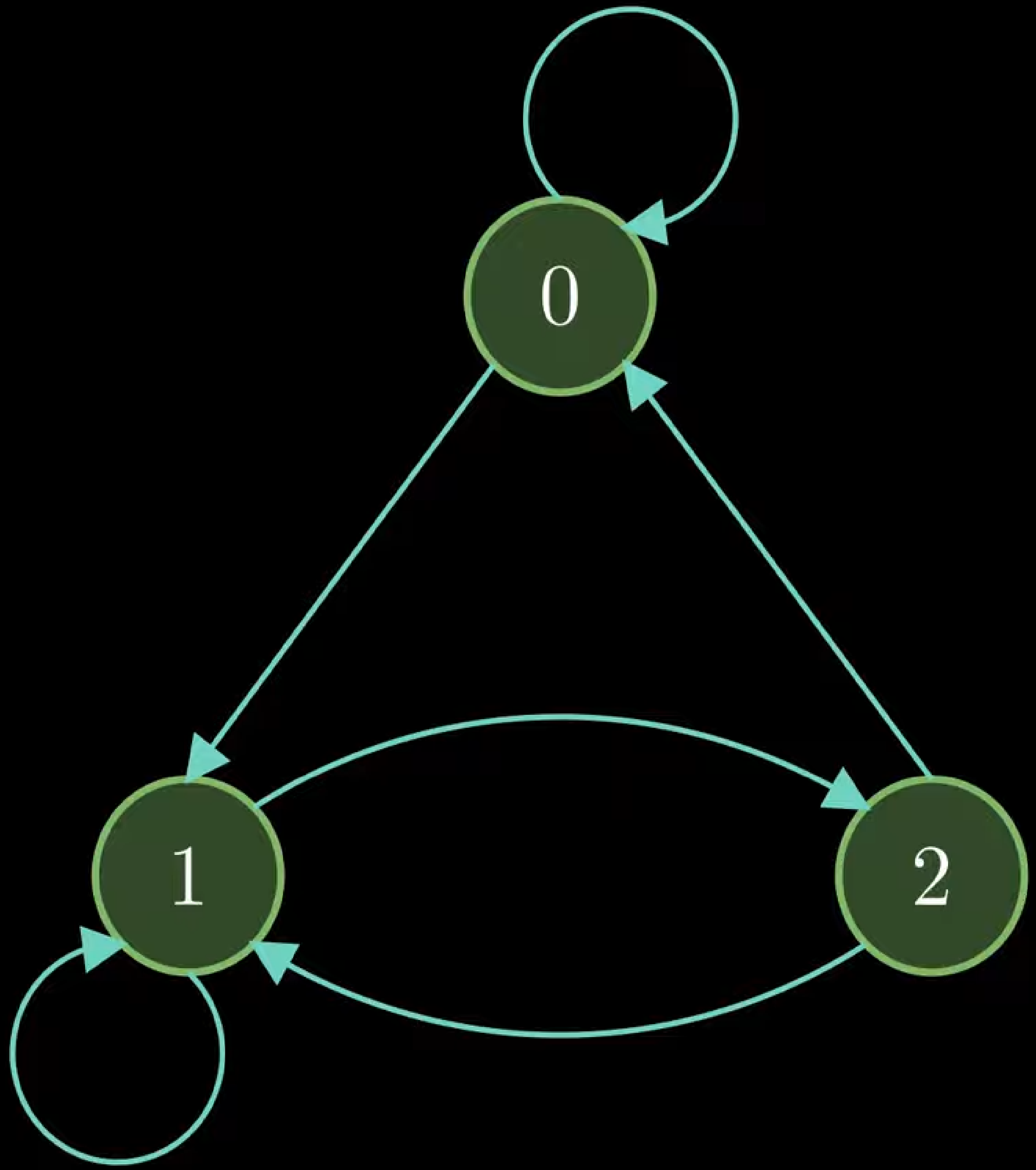
$$A = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$\pi_0 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

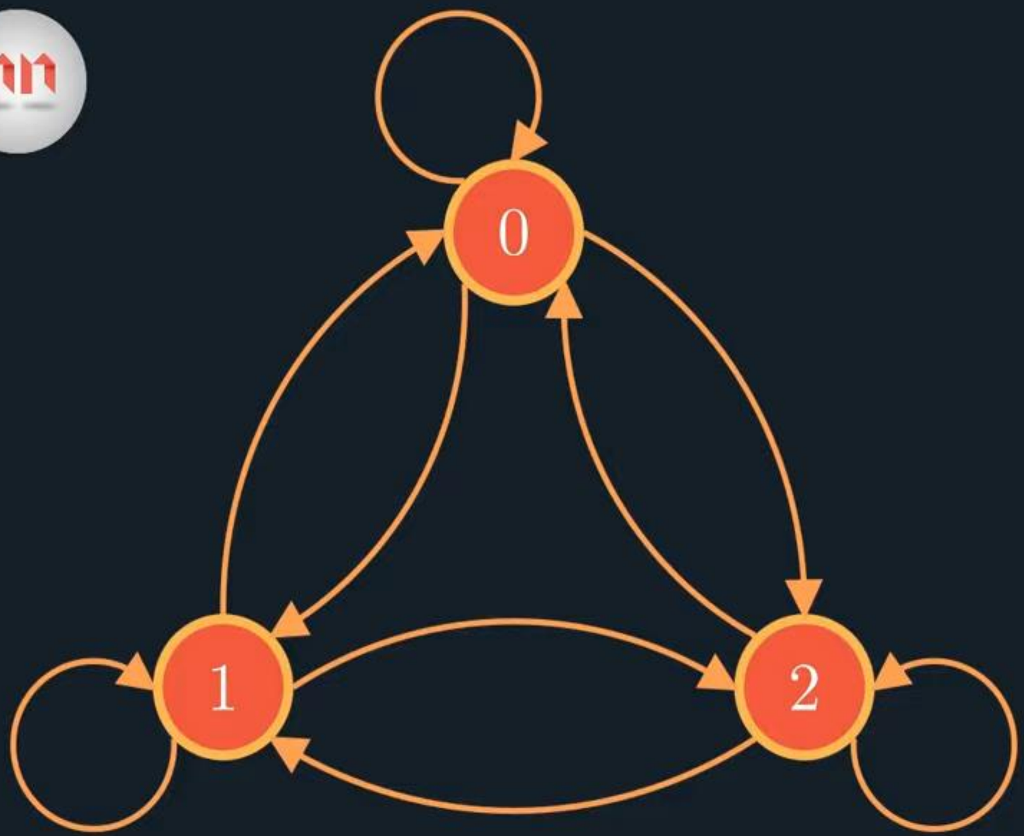
$$\pi = \begin{bmatrix} 0.35211 & 0.21127 & 0.43662 \end{bmatrix}$$

$$0.35191 \quad 0.21245 \quad 0.43564$$


0
1
2
0
Transient State
1
2
Recurrent State



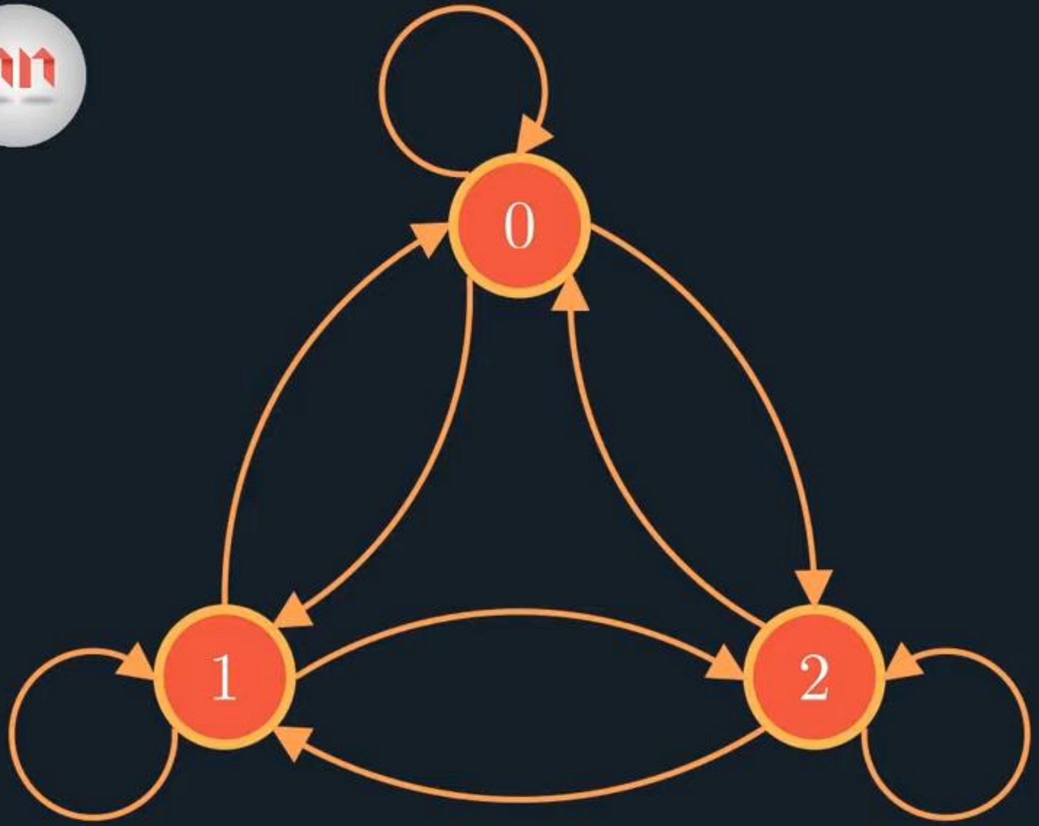
Irreducible



Probability of reaching state j from state i after exactly n steps

$$P_{ij}(n)$$

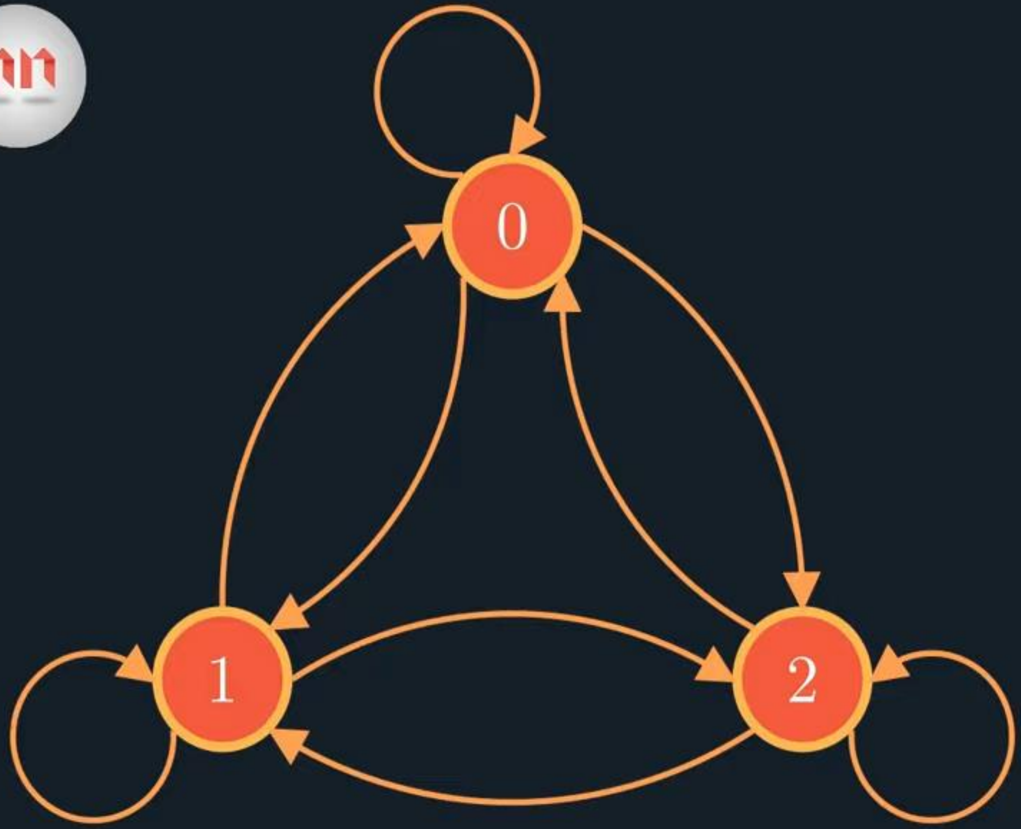
$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$



$$A = \begin{bmatrix} 0.5 & 0.2 & \boxed{0.3} \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$

Probability of reaching state 2 from state 0 after exactly 1 step

$$P_{02}(1)$$



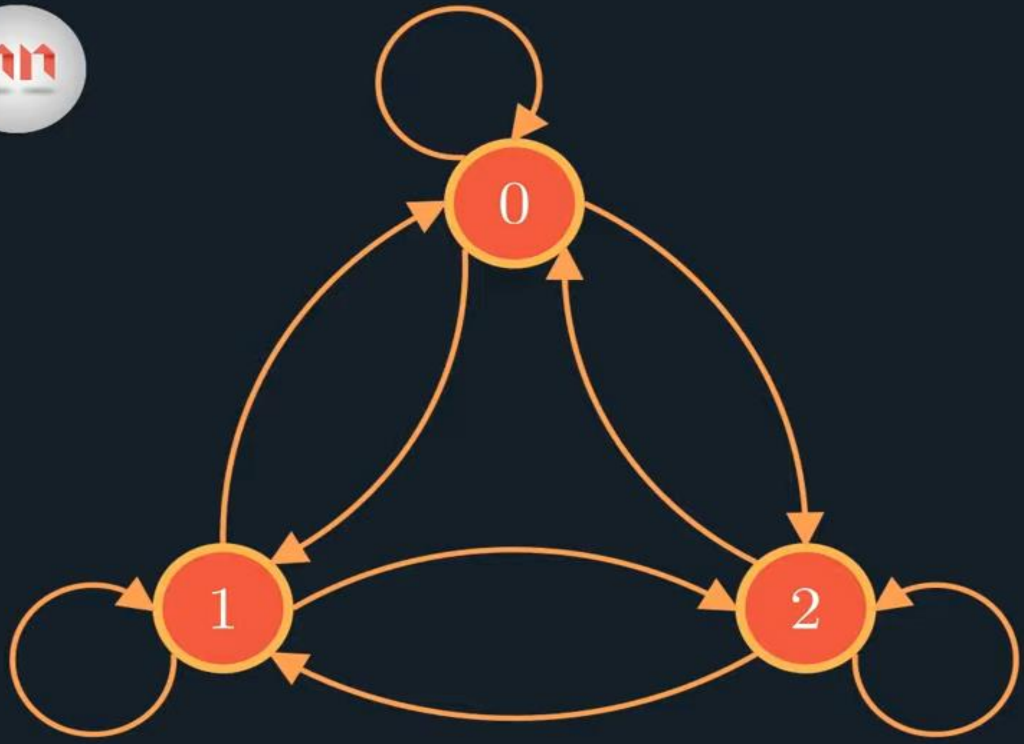
Probability of reaching state 2 from state 0 after exactly 2 steps

$$P_{02}(2)$$

$$A_{01} \times A_{12} + A_{00} \times A_{02} + A_{02} \times A_{22}$$

$$0.2 \times 0.2 + 0.5 \times 0.3 + 0.3 \times 0.1$$

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$

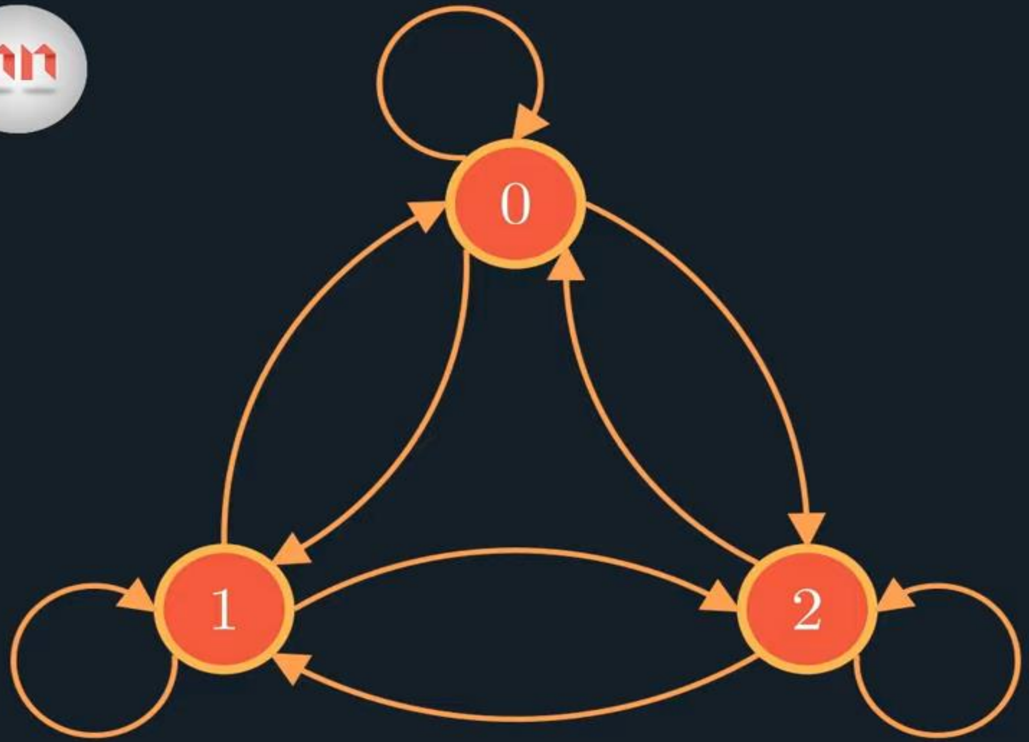


$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix}$$

Probability of reaching state 2 from state 0 after exactly 2 steps

$$P_{02}(2)$$

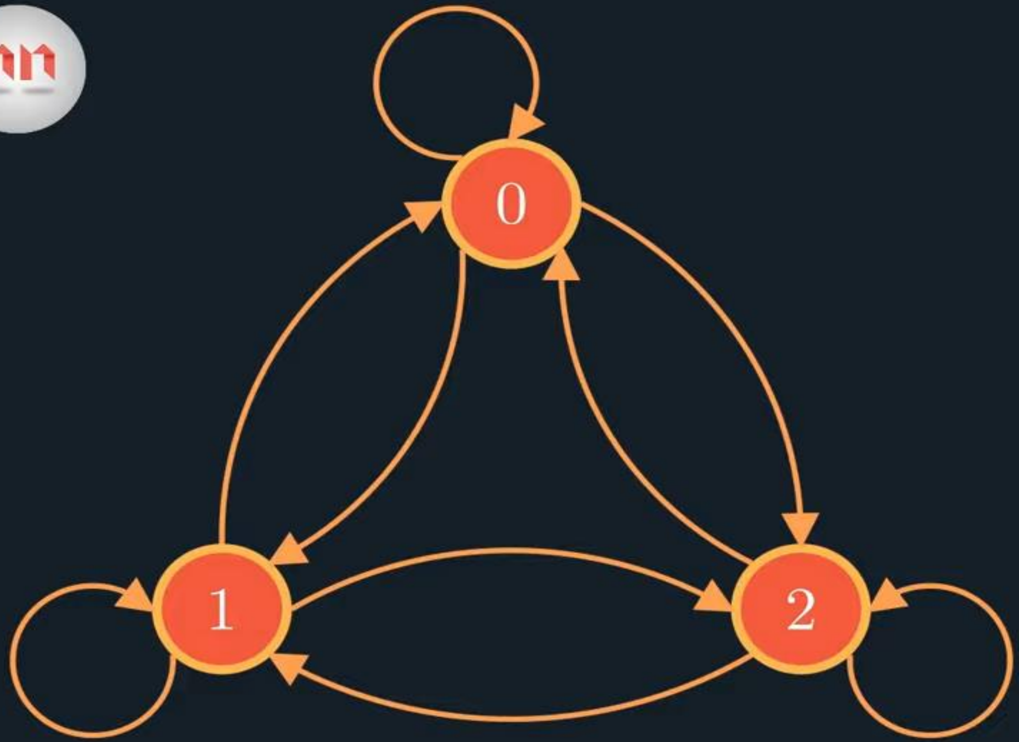
$$\begin{bmatrix} A_{00} & A_{01} & A_{02} \end{bmatrix} \times \begin{bmatrix} A_{02} \\ A_{12} \\ A_{22} \end{bmatrix}$$



$$P_{02}(2) = \begin{bmatrix} A_{00} & A_{01} & A_{02} \end{bmatrix} \times \begin{bmatrix} A_{02} \\ A_{12} \\ A_{22} \end{bmatrix}$$

$$P_{10}(2) = \begin{bmatrix} A_{10} & A_{11} & A_{12} \end{bmatrix} \times \begin{bmatrix} A_{00} \\ A_{10} \\ A_{20} \end{bmatrix}$$

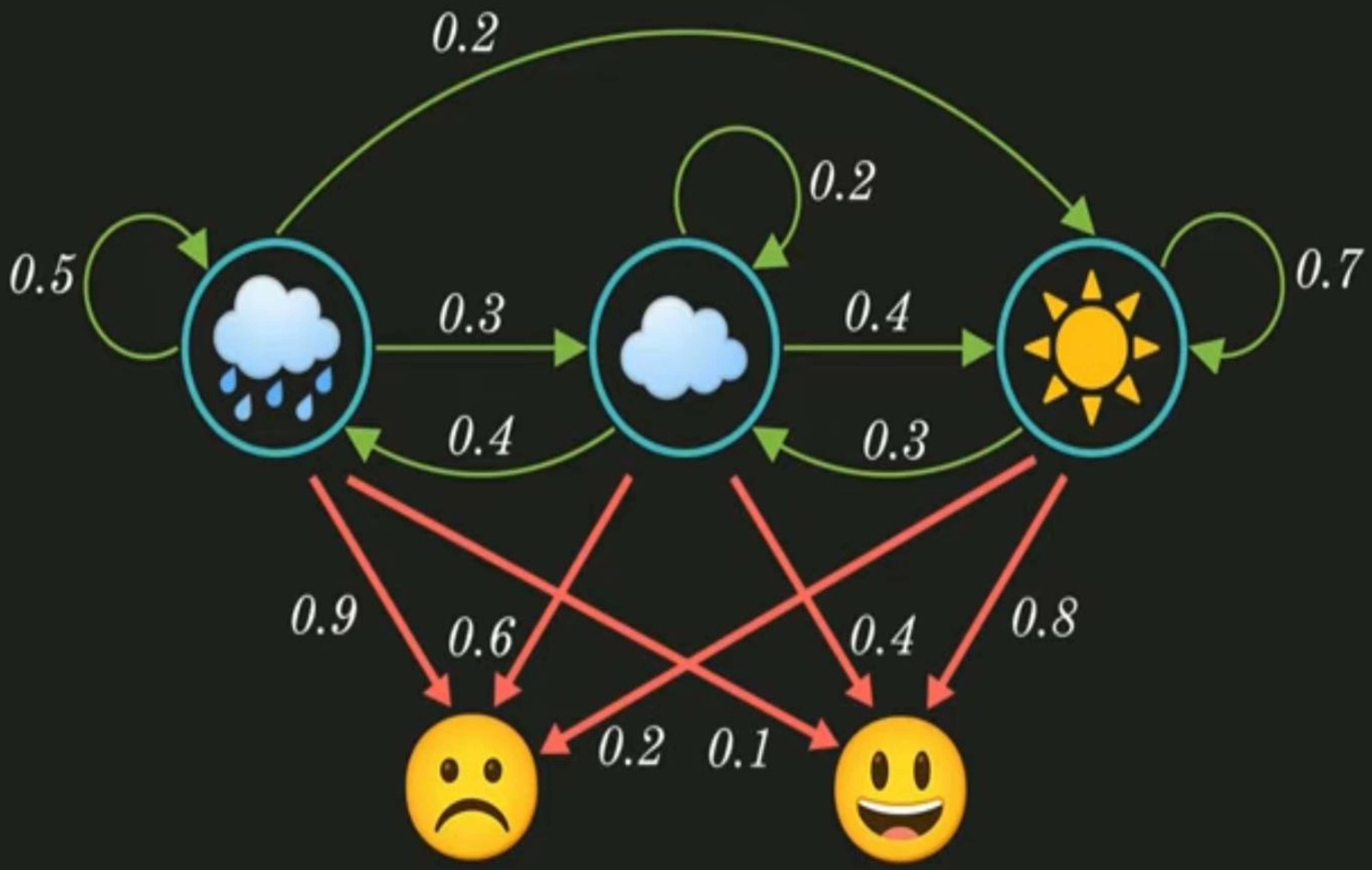
$$A^2 = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} \times \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} = \begin{bmatrix} 0.40 & 0.38 & \boxed{0.22} \\ 0.44 & 0.32 & 0.24 \\ 0.54 & 0.26 & 0.20 \end{bmatrix}$$

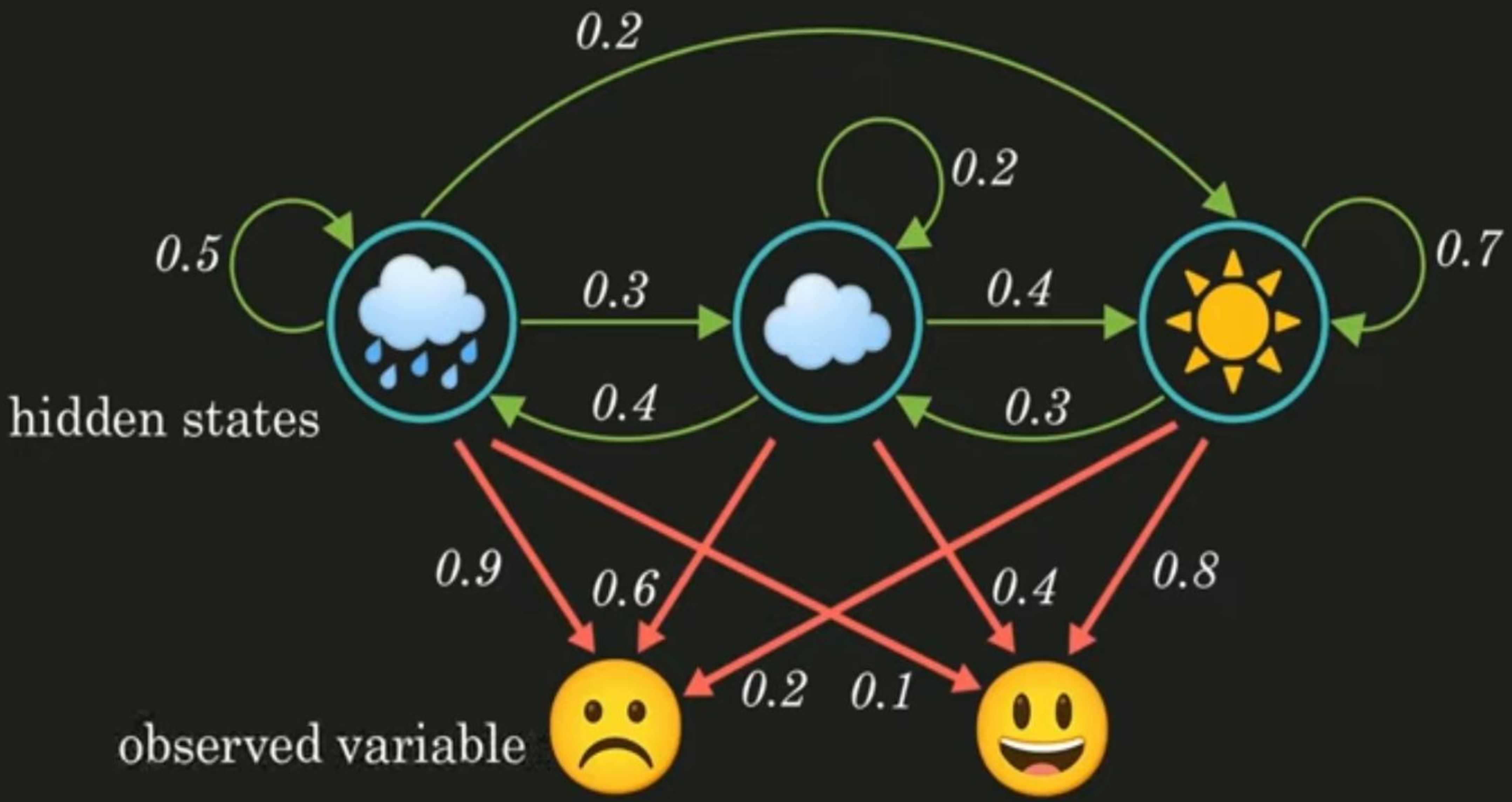


$$P_{ij}(2) = A^2_{ij}$$

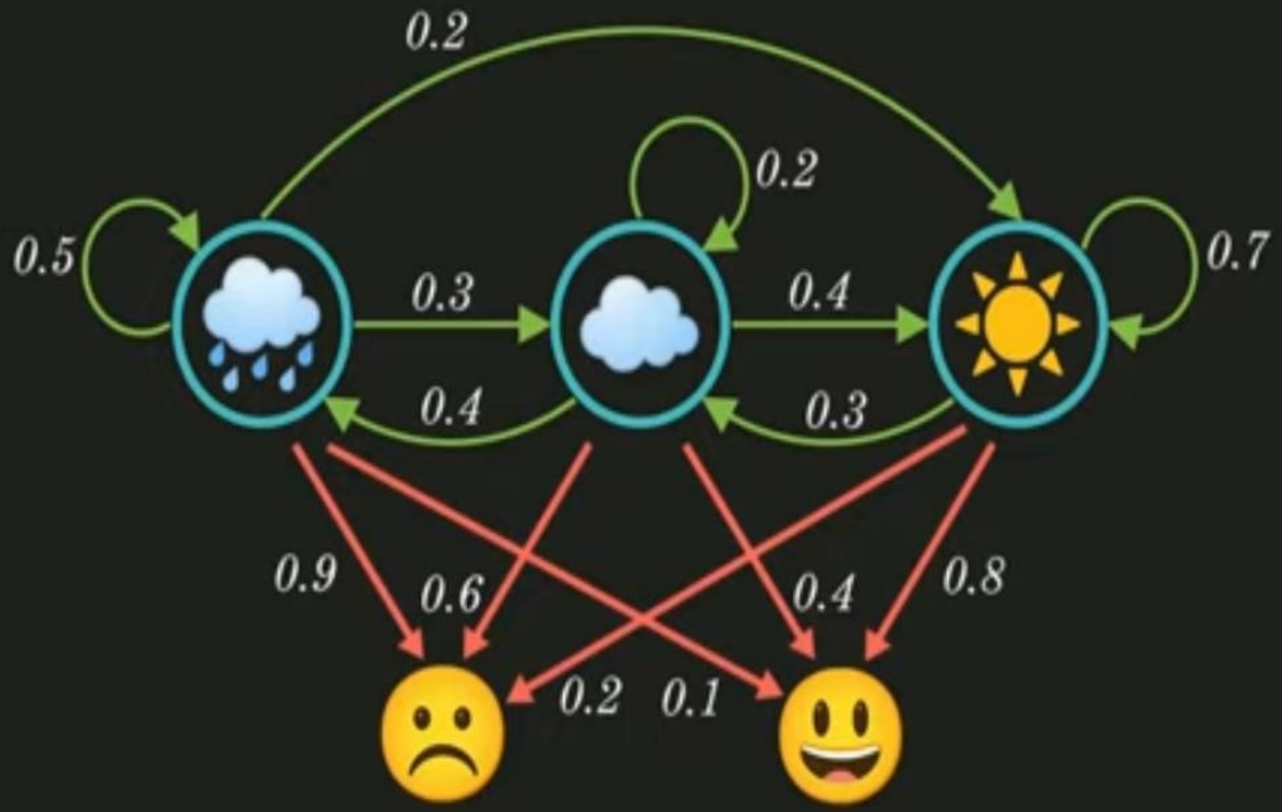
$$P_{ij}(n) = A^n_{ij}$$

$$A^2 = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} \times \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{bmatrix} = \begin{bmatrix} 0.40 & 0.38 & 0.22 \\ 0.44 & 0.32 & 0.24 \\ 0.54 & 0.26 & 0.20 \end{bmatrix}$$


0.2
0.2
0.5
0.7
0.3
0.4
0.4
0.3
0.9
0.6
0.4
0.8
0.2
0.1

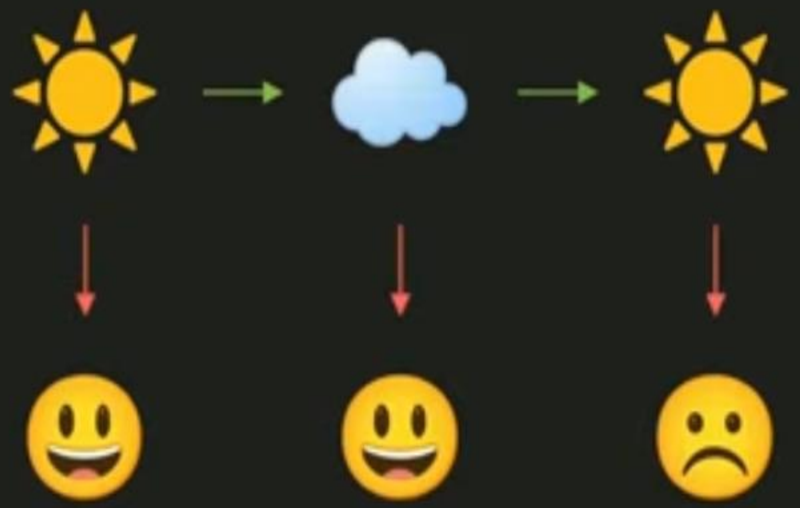


0.2
0.2
0.5
0.7
0.3
0.4
0.4
0.3
hidden states
0.9
0.6
0.4
0.8
0.2
0.1
observed variable



	🌧️	☁️	☀️
🌧️	0.5	0.3	0.2
☁️	0.4	0.2	0.4
☀️	0.0	0.3	0.7

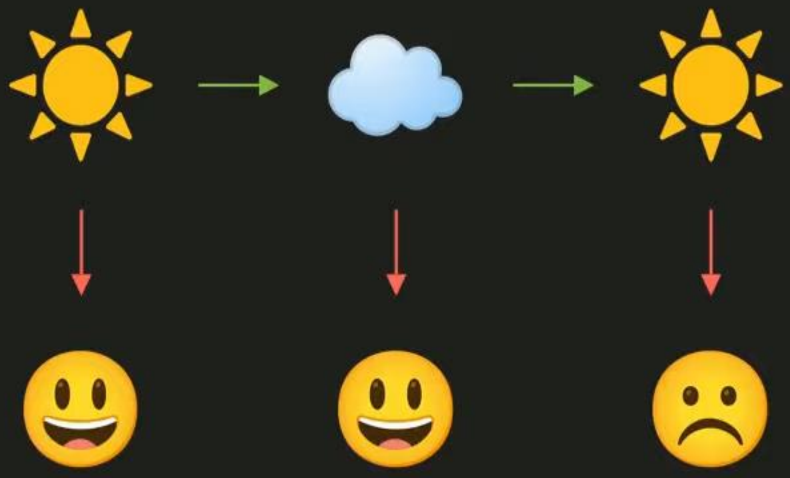
☹️	😃
0.9	0.1
0.6	0.4
0.2	0.8



$$\begin{array}{c|ccc} & 🌧 & ☁ & ☀ \\ \hline 🌧 & 0.5 & 0.3 & 0.2 \\ ☁ & 0.4 & 0.2 & 0.4 \\ ☀ & 0.0 & 0.3 & 0.7 \end{array} \qquad \begin{array}{cc} ☹ & 😃 \\ \hline 0.9 & 0.1 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{array}$$

What is the probability of occurring this sequence?

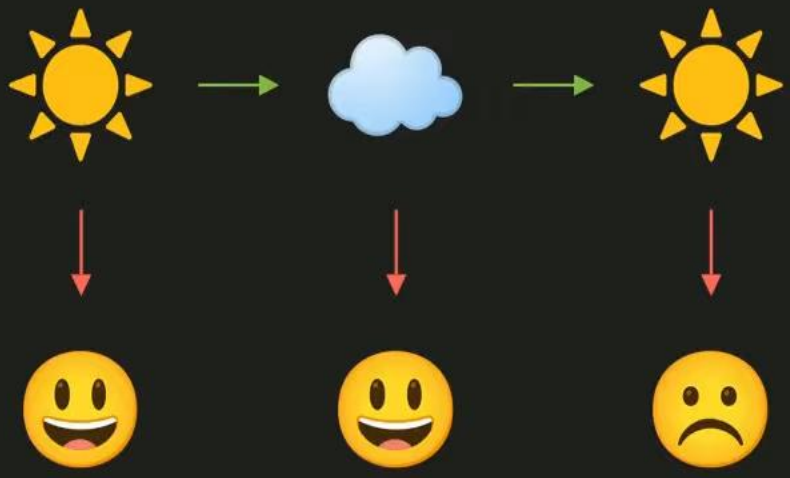
$P(Y =$ 😃 😃 ☹ $, X =$ ☀ ☁ ☀ $)$



	🌧️	☁️	☀️
🌧️	0.5	0.3	0.2
☁️	0.4	0.2	0.4
☀️	0.0	0.3	0.7

	☹️	😃
🌧️	0.9	0.1
☁️	0.6	0.4
☀️	0.2	0.8

$P(X_1 = \text{☀️})$	$P(Y_1 = \text{😃} \mid X_1 = \text{☀️})$
?	0.8
$P(X_2 = \text{☁️} \mid X_1 = \text{☀️})$	$P(Y_2 = \text{😃} \mid X_2 = \text{☁️})$
0.3	0.4
$P(X_3 = \text{☀️} \mid X_2 = \text{☁️})$	$P(Y_3 = \text{☹️} \mid X_3 = \text{☀️})$
0.4	0.2



	🌧️	☁️	☀️
🌧️	0.5	0.3	0.2
☁️	0.4	0.2	0.4
☀️	0.0	0.3	0.7

	☹️	😃
🌧️	0.9	0.1
☁️	0.6	0.4
☀️	0.2	0.8

$P(X_1 = \text{☀️})$ $P(Y_1 = \text{😃} \mid X_1 = \text{☀️})$

0.509 0.8

$P(X_2 = \text{☁️} \mid X_1 = \text{☀️})$ $P(Y_2 = \text{😃} \mid X_2 = \text{☁️})$

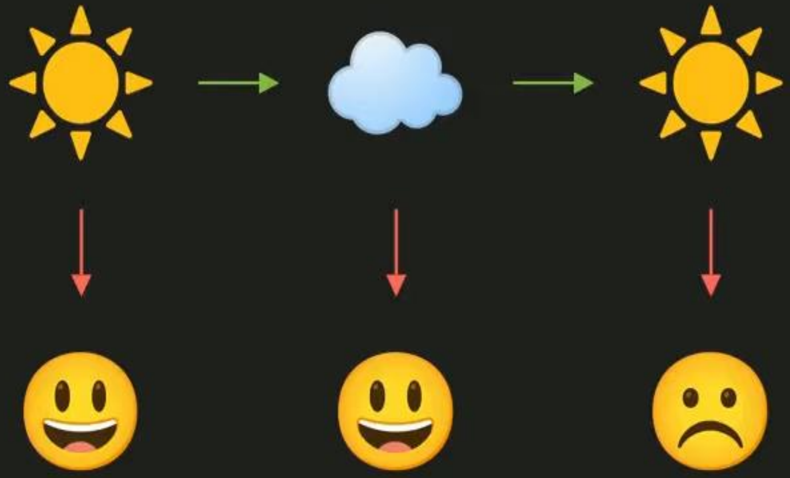
0.3 0.4

$P(X_3 = \text{☀️} \mid X_2 = \text{☁️})$ $P(Y_3 = \text{☹️} \mid X_3 = \text{☀️})$

0.4 0.2

$$\pi A = \pi$$

$$\pi = [0.218, 0.273, 0.509]$$

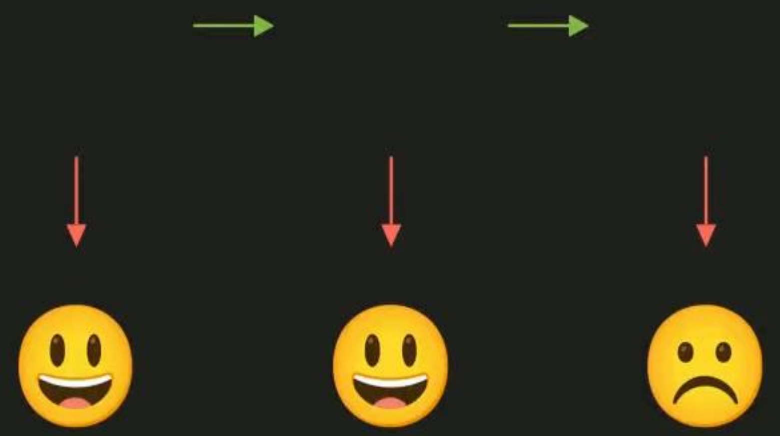


$$
\begin{array}{c|ccc}
 & 🌧️ & ☁️ & ☀️ \\
\hline
🌧️ & 0.5 & 0.3 & 0.2 \\
☁️ & 0.4 & 0.2 & 0.4 \\
☀️ & 0.0 & 0.3 & 0.7
\end{array}
\qquad
\begin{array}{cc}
☹️ & 😃 \\
\hline
0.9 & 0.1 \\
0.6 & 0.4 \\
0.2 & 0.8
\end{array}
$$

$$\underset{0.509}{P(X_1 = \text{☀️})} \quad \underset{0.8}{P(Y_1 = \text{😃} \mid X_1 = \text{☀️})}$$

$$\underset{0.3}{P(X_2 = \text{☁️} \mid X_1 = \text{☀️})} \quad \underset{0.4}{P(Y_2 = \text{😃} \mid X_2 = \text{☁️})} \qquad 0.509 * 0.8 * 0.3 * 0.4 * 0.4 * 0.2$$

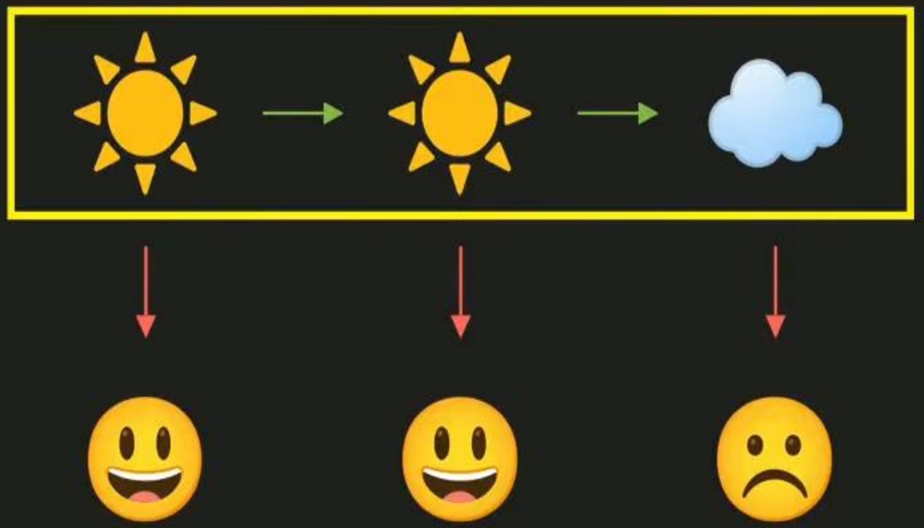
$$\underset{0.4}{P(X_3 = \text{☀️} \mid X_2 = \text{☁️})} \quad \underset{0.2}{P(Y_3 = \text{☹️} \mid X_3 = \text{☀️})}$$



	🌧️	☁️	☀️
🌧️	0.5	0.3	0.2
☁️	0.4	0.2	0.4
☀️	0.0	0.3	0.7

	☹️	😃
🌧️	0.9	0.1
☁️	0.6	0.4
☀️	0.2	0.8

What is the most likely weather sequence for the observed mood sequence?

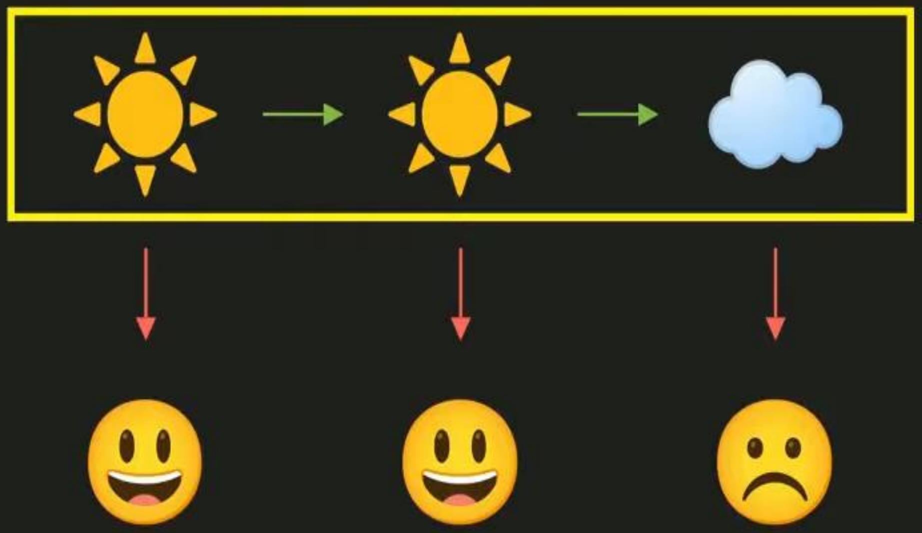


	🌧️	☁️	☀️
🌧️	0.5	0.3	0.2
☁️	0.4	0.2	0.4
☀️	0.0	0.3	0.7

☹️	😃
0.9	0.1
0.6	0.4
0.2	0.8

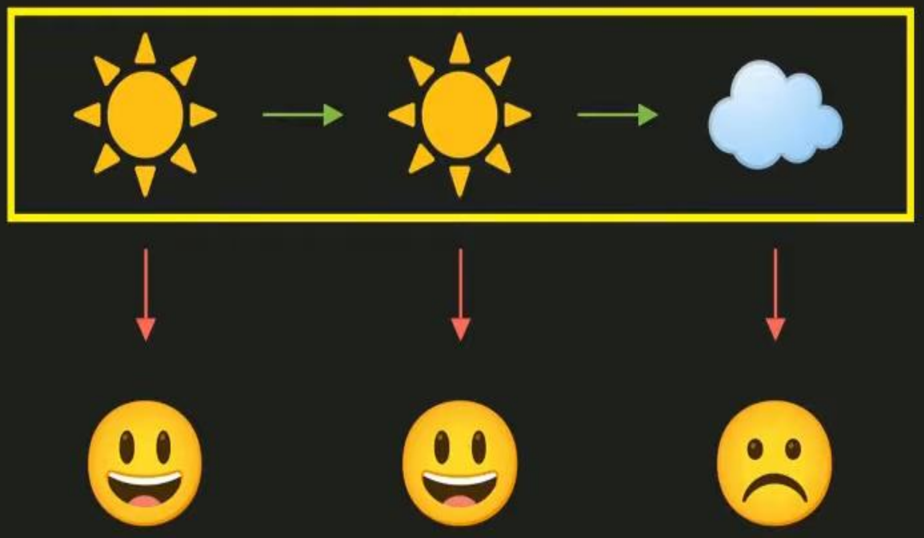
$P(Y =$ 😃 😃 ☹️ $, X =$ ☀️ ☀️ ☁️ $)$

$= 0.04105$



$$
\begin{array}{c|ccc}
 & \text{🌧️} & \text{☁️} & \text{☀️} \\
\hline
\text{🌧️} & 0.5 & 0.3 & 0.2 \\
\text{☁️} & 0.4 & 0.2 & 0.4 \\
\text{☀️} & 0.0 & 0.3 & 0.7
\end{array}
\qquad
\begin{array}{cc}
\text{☹️} & \text{😃} \\
\hline
0.9 & 0.1 \\
0.6 & 0.4 \\
0.2 & 0.8
\end{array}
$$

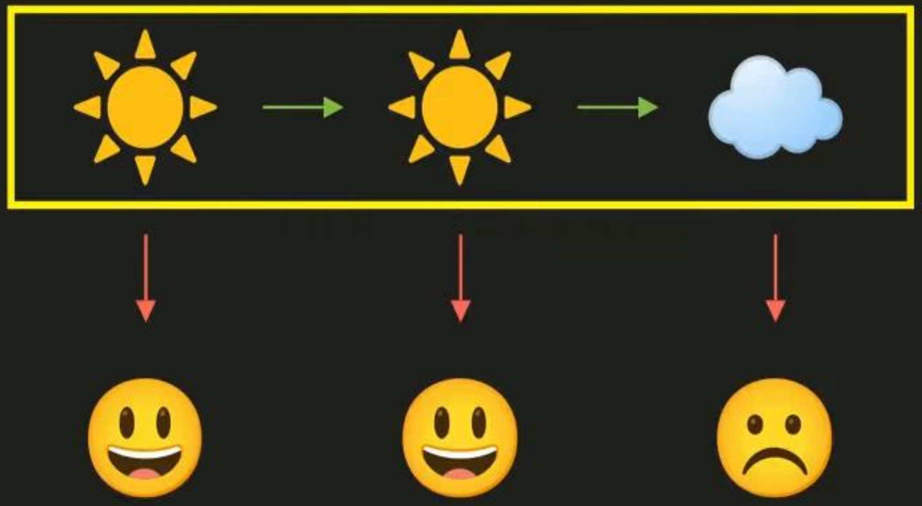
$$\underset{X=X_1,X_2,\ldots X_n}{\arg\max} \; P(X = X_1, X_2, \ldots X_n \mid Y = Y_1, Y_2, \ldots Y_n)$$



$$
\begin{array}{cc}
 & \begin{array}{ccc} 🌧️ & ☁️ & ☀️ \end{array} \\
\begin{array}{c} 🌧️ \\ ☁️ \\ ☀️ \end{array} &
\begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.4 & 0.2 & 0.4 \\ 0.0 & 0.3 & 0.7 \end{bmatrix}
\end{array}
\qquad
\begin{array}{c}
\begin{array}{cc} ☹️ & 😃 \end{array} \\
\begin{bmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}
\end{array}
$$

$$\underset{X=X_1,X_2,\ldots X_n}{\arg\max} \frac{P(Y|X)P(X)}{P(Y)}$$

$$P(Y|X) = P(Y_1|X_1) * P(Y_2|X_2) * \ldots P(Y_n|X_n)$$



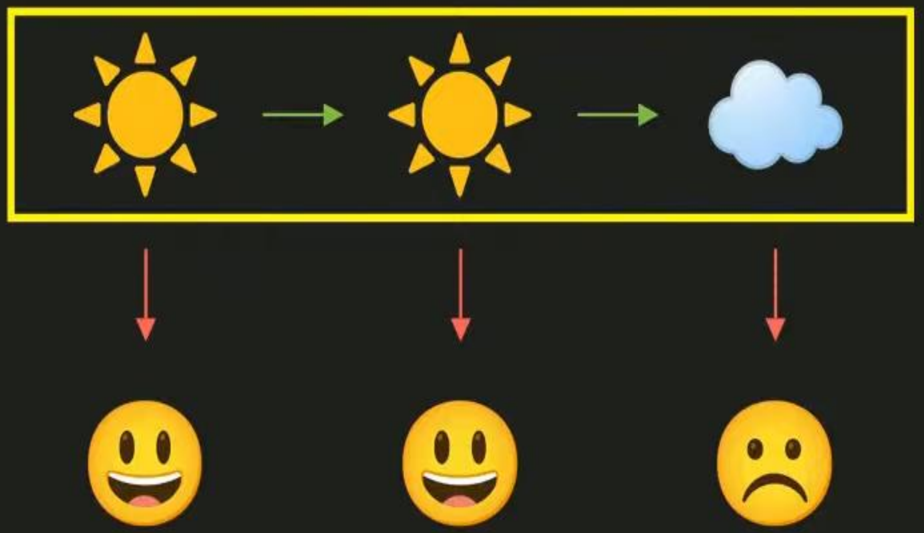
	🌧️	☁️	☀️
🌧️	0.5	0.3	0.2
☁️	0.4	0.2	0.4
☀️	0.0	0.3	0.7

	☹️	😃
🌧️	0.9	0.1
☁️	0.6	0.4
☀️	0.2	0.8

$$\underset{X=X_1,X_2,\ldots X_n}{\arg\max} \frac{P(Y|X)P(X)}{P(Y)}$$

$$P(Y|X) = \prod P(Y_i \mid X_i)$$

$$P(X) = \prod P(X_i \mid X_{i-1})$$



$$
\begin{array}{c|ccc}
 & 🌧 & ☁ & ☀ \\
\hline
🌧 & 0.5 & 0.3 & 0.2 \\
☁ & 0.4 & 0.2 & 0.4 \\
☀ & 0.0 & 0.3 & 0.7
\end{array}
\qquad
\begin{array}{cc}
☹ & 😃 \\
\hline
0.9 & 0.1 \\
0.6 & 0.4 \\
0.2 & 0.8
\end{array}
$$

$$\underset{X=X_1,X_2,\ldots X_n}{\arg\max} \prod P(Y_i \mid X_i)\, P(X_i \mid X_{i-1})$$

AUTO INSURANCE RISK

Based on company data, a motorist that is currently high risk has a 60% chance of being denoted high risk again when the policy renews and a 40% chance of being moved to low risk.

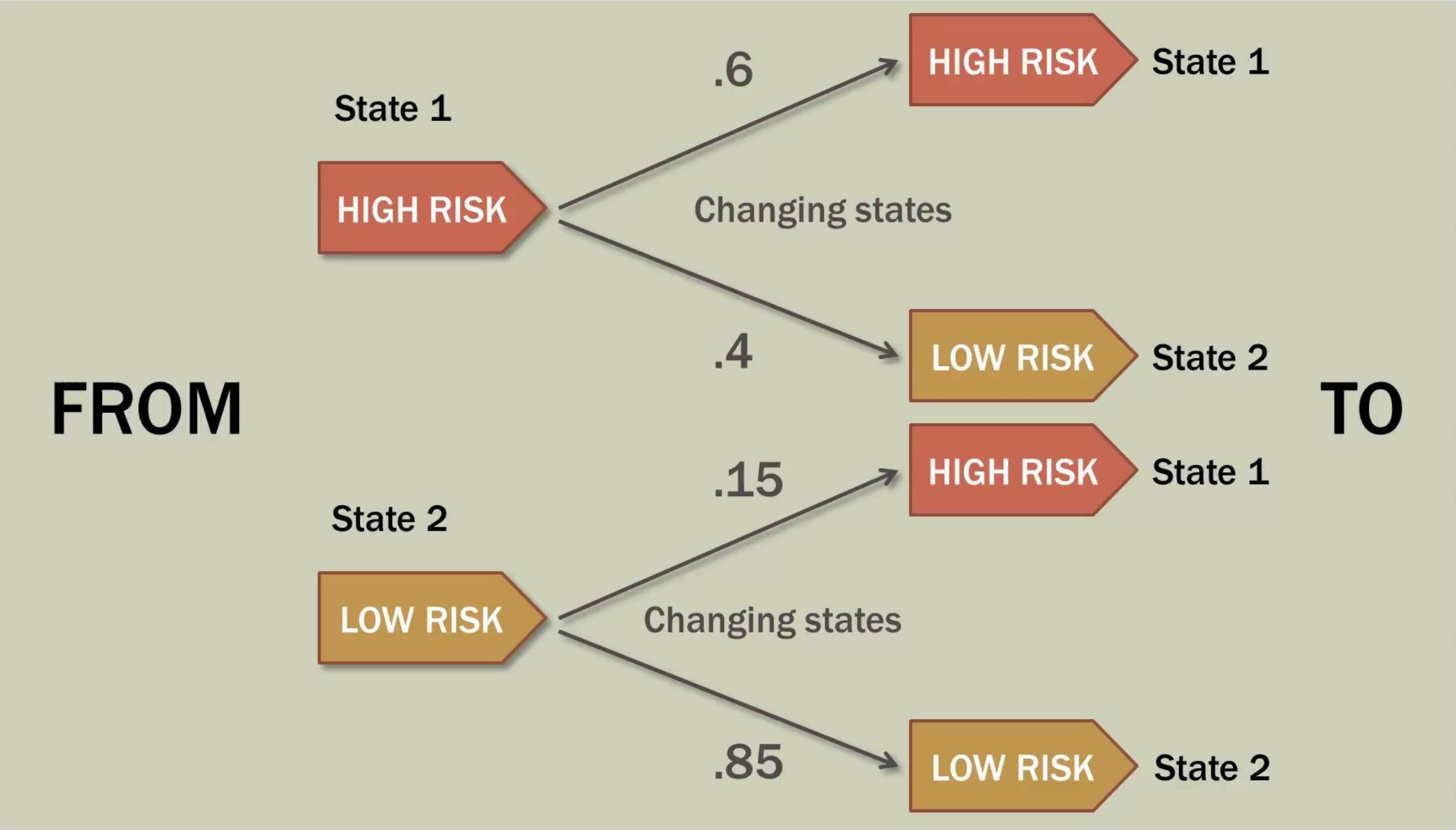
A low risk driver on the other hand has a 15% chance of moving to the high risk category and an 85% chance of remaining low risk.

Task:

Set up a probability tree, transition diagram, and transition matrix for our process.

"STATE"

- A "state" is simply the category (in this problem) a motorist can be in at any given time.
- Each driver is either in the HIGH RISK state or the LOW RISK state.
- Every driver is in a state and no drivers are in both states.
- Drivers can move between states or return to the same state when their policy is renewed.


FROM
TO
State 1
HIGH RISK
.6
HIGH RISK
State 1
Changing states
.4
LOW RISK
State 2
State 2
LOW RISK
.15
HIGH RISK
State 1
Changing states
.85
LOW RISK
State 2

Markov chain key features:

A sequence of trials of an experiment is a Markov chain if:

1. the outcome of each experiment is one of a set of discrete states;
2. the outcome of an experiment depends only on the present state, and not on any past states.

Transition Matrix : contains all the conditional probabilities of the Markov chain

$$p_{ij} = \text{Prob(State } n+1 \text{ is } S_i \mid \text{State } n \text{ is } S_j),$$

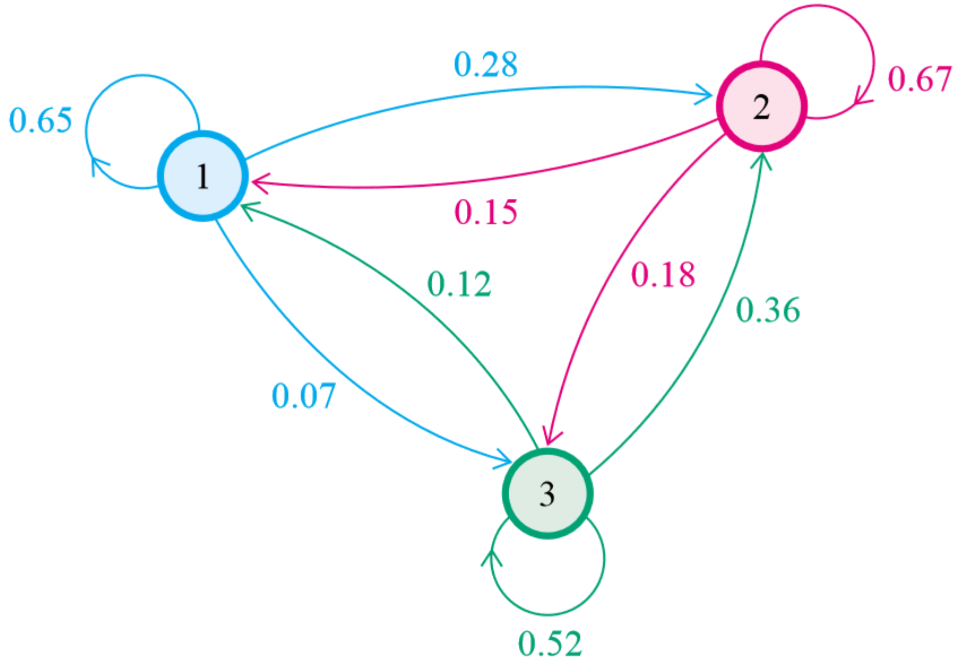
$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & & \ddots & \\ p_{r1} & & & p_{rr} \end{bmatrix}$$

Where Pij is the conditional probability of being in state Si at step n+1 given that the process was in state Sj at step n.

Example:

$$\text{State } n+1 \left\{ \begin{array}{c|ccc|} & \overbrace{S_1 \quad S_2 \quad S_3}^{\text{State } n} & & \\ S_1 & p_{11} & p_{12} & p_{13} \\ S_2 & p_{21} & p_{22} & p_{23} \\ S_3 & p_{31} & p_{32} & p_{33} \end{array} \right.$$

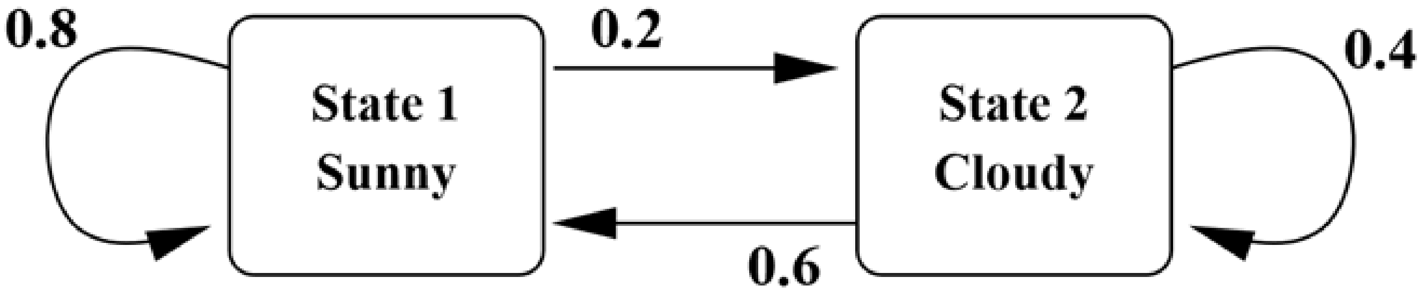
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} \end{array} = P.$$



Example: Sunny or Cloudy. A meteorologist studying the weather in a region decides to classify each day as simply *sunny* or *cloudy*. After analyzing several years of weather records, he finds:

- the day after a sunny day is sunny 80% of the time, and cloudy 20% of the time; and
- the day after a cloudy day is sunny 60% of the time, and cloudy 40% of the time.

We can setup up a Markov chain to model this process. There are just two states: S_1 = *sunny*, and S_2 = *cloudy*. The transition diagram is



and the transition matrix is

$$P = \begin{bmatrix} 0.8 & \mathbf{0.2} \\ \mathbf{0.6} & 0.4 \end{bmatrix}.$$

Transition matrix features

- It is square, since all possible states must be used both as rows and as columns.
- All entries are between 0 and 1, because all entries represent probabilities.
- The sum of the entries in any row must be 1, since the numbers in the row give the probability of changing from the state at the left to one of the states indicated across the top.

special cases of Markov chains

- **regular Markov chains:**

A Markov chain is a regular Markov chain if some power of the transition matrix has only positive entries. That is, if we define the (i; j) entry of P^n to be p^n_{ij} , then the Markov chain is regular if there is some n such that $p^n_{ij} > 0$ for all (i,j).

- **absorbing Markov chains:**

A state S_k of a Markov chain is called an absorbing state if, once the Markov chains enters the state, it remains there forever.

A Markov chain is called an absorbing chain if

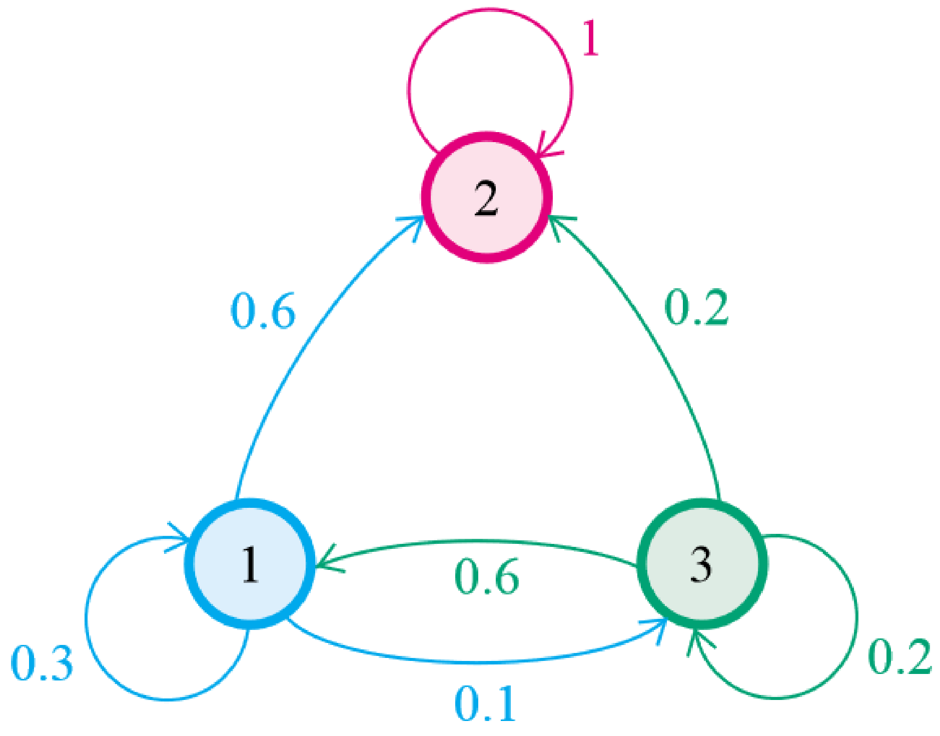
❑ It has at least one absorbing state.

❑ For every state in the chain, the probability of reaching an absorbing state in a finite number of steps is nonzero.

Examples : Absorbing

State 2 is absorbing

Pii = 1 → P22 = 1



$$
\begin{array}{cc}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} = P.
\end{array}
$$

http://www.math.bas.bg/~jeni/markov123.pdf

Some applications:

- **Natural Language Processing:** HMMs are used in applications such as part-of-speech tagging, named entity recognition, and text-to-speech synthesis. HMMs can model the probabilistic transitions between different linguistic states, helping to analyze and generate natural language.

- **Handwriting Recognition:** HMMs have been employed in handwriting recognition systems. They can model the sequence of pen strokes or features extracted from handwriting, allowing for the recognition of handwritten characters or words.

- **Gesture Recognition:** HMMs can be used to recognize and interpret human gestures in applications like sign language recognition, motion capture, and gesture-based interfaces. HMMs can capture the temporal dynamics of gesture sequences and classify them into predefined classes.

Some applications:

- **Bioinformatics:** HMMs have applications in bioinformatics, such as gene finding, protein structure prediction, and sequence alignment. HMMs can model the probabilistic relationships between DNA or protein sequences, aiding in the identification of functional elements or similarities.

- **Financial Modeling:** HMMs are utilized in financial modeling for tasks such as predicting stock prices, analyzing market trends, and risk assessment. HMMs can capture hidden states representing different market conditions and their transitions over time.

- **Computer Vision:** HMMs find applications in computer vision tasks like object tracking, activity recognition, and video analysis. HMMs can model the temporal dynamics and probabilistic relationships between observed visual features, enabling efficient analysis of video data.