计算机应用数学-2024 秋

Lecture 1.3: 随机游走与马尔可夫链

2024.9.27

Lecturer: 宋骐 *Scribe:* 王亦涵, 赵奇, 王馨语, 郑岭

1 Introduction

1.1 Definition

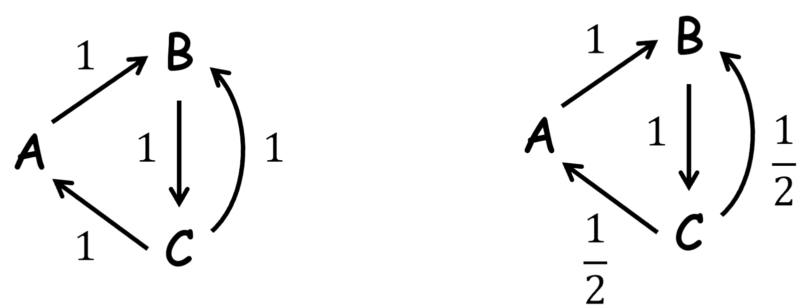
Random walk on a directed graph: a sequence of vertices generated from a start vertex by randomly selecting an incident edge, traversing the edge to a new vertex, and repeating the process.

Formally:

$$p(t)P = p(t+1)$$

where $p(t)$ is a row vector with a component for each vertex specifying the probabilities mass of the vertex at time t, P is the so called transition matrix, and $P_{i,j}$ is the probability of the walk at vertex i selecting j.

Example 1.3.1

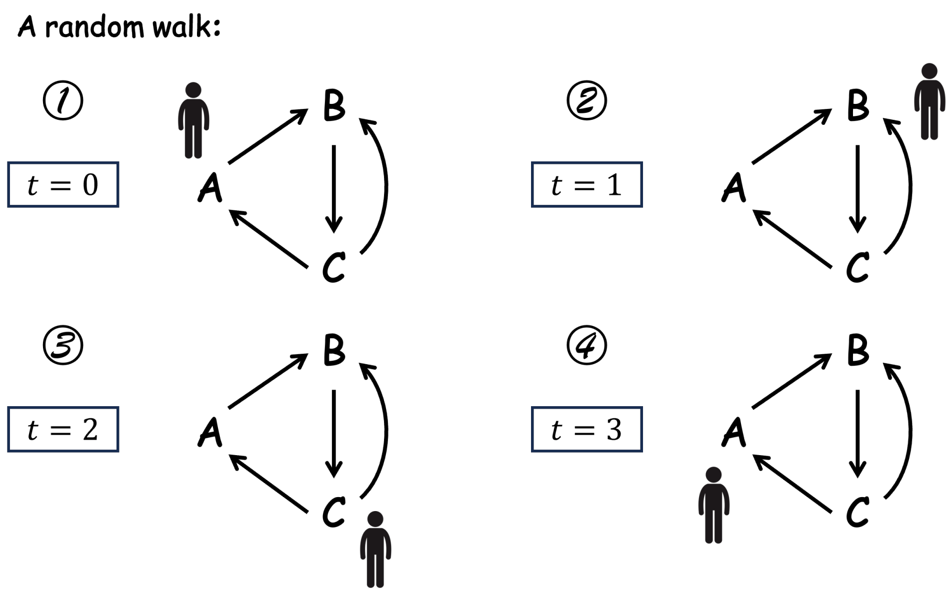


Adjacency Matrix A

$$\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array}$$

Transition Matrix P

$$\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array}$$



Markov chain.

A finite set of states.

p_{xy}: transition probability of going from state x to y, $\sum_y p_{xy} = 1$.

A Markov chain can be represented as a directed graph with weight p_{xy} from vertex x to y.

Random walk	Markov chain
Graph	Stochastic process
Vertex	State
Strongly connected	persistent
Aperiodic	Aperiodic
Strongly connected and aperiodic	Ergodic
Undirected graph	Time reversible

We will introduce the following in this section.

* Example: PageRank and Markov Decision Process.
* Stationary distribution.
* Convergence.
* Markov Process.

1.2 Examples

PageRank.

Consider Web as a graph: each webpage is a vertex, the hyperlinks are edges.

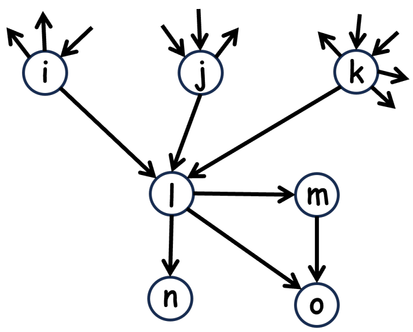
Goal: rank webpages based on their importance.

Insight.

A webpage is more important if it has more links.

Consider in-coming links as votes, famous websites have more incoming links.

Moreover, links from important pages count more.



$$r_L = \frac{r_i}{3} + \frac{r_j}{2} + \frac{r_k}{4}$$

$$r_j = \sum_{i \to j} \frac{r_i}{d_i} \qquad (*)$$

Stochastic adjacency matrix

d_i is the degree of node i.

if $i \Rightarrow j$, $M_{ji} = \dfrac{1}{d_i}$

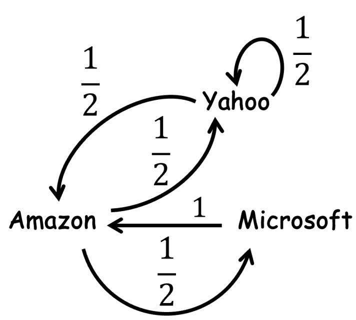
Rank vector

r_i is the importance score of page i

$(*)$ can be rewritten as

$r = M \cdot r$

Example 1.3.2



$$M = \begin{array}{c} \\ \mathbf{Y} \\ \mathbf{A} \\ \mathbf{M} \end{array} \!\!\! \begin{array}{c} \begin{array}{ccc} \mathbf{Y} & \mathbf{A} & \mathbf{M} \end{array} \\ \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1 \\ 0 & 1/2 & 0 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{A} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

1^{st} iteration

$$\begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{1}{6} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$\begin{aligned} PR(Y)^1 &= \frac{1}{2}PR(Y)^0 + \frac{1}{2}PR(A)^0 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3} \\ PR(A)^1 &= \frac{1}{2}PR(Y)^0 + 1 \cdot PR(M)^0 = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} = \frac{3}{6} \\ PR(M)^1 &= \frac{1}{2}PR(A)^0 = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6} \end{aligned}$$

2^{nd} iteration

$$\begin{bmatrix} \frac{5}{12} \\ \frac{1}{3} \\ \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \\ \frac{1}{6} \end{bmatrix}$$

...

convergence

$$\begin{bmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}$$

<u>Markov Process</u> (Markov Decision Process).

$$\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathbb{P}, \gamma$$

$\mathcal{S}$: set of states.

$\mathcal{A}$: set of actions.

$\mathcal{R}$: reward $r(s, a)$ under state s take action a.

$\mathbb{P}$: transition probability $P(s'|s,a)$ under state s take action a, next state s'.

γ: discount factor.

MDP:

$t = 0$ initial state $s_0 \sim p(s_0)$

For $t = 0$ to end:

take action a_t

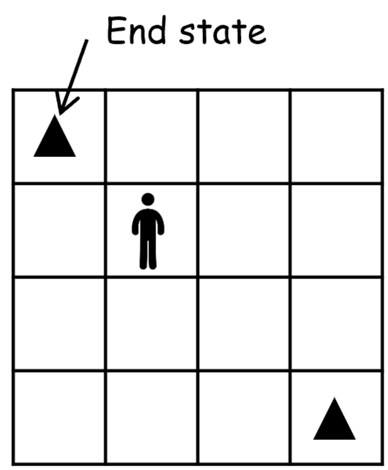
get reward $r_t \sim R(\cdot|s_t, a_t)$

get next state $s_{t+1} \sim P(\cdot|s_t, a_t)$

agent get reward $r_t \;\; \& \;\; s_{t+1}$

object: maximize long-term reward (accumulated) $\sum_{t\geq 0} D^t r_t$.

Example 1.3.3



Actions = {left, right, up, down}

reward to empty block $\rightarrow$ use minimized actions to achieve end state.

2 Stationary Distribution

Let p_t be the probability distribution after t steps of a random walk. Define the long-term average probability distribution a_t by

$$a_t = \frac{1}{t}\left(p_0 + p_1 + \cdots + p_{t-1}\right)$$

The fundamental theorem of Markov Chains:

For a connected Markov Chain, at converges to a limit probability vector x, which satisfies

$$XP = x; \sum_i x_i = 1 \Rightarrow X[P - I, 1] = [0, 1]$$

Lemma 1.3.1. Let P be the transition probability matrix for a connected Markov Chain. The $n \times (n + 1)$ matrix $A = [P - I, 1]$ obtained by augmenting the matrix $P - I$ with an additional column of ones has rank n.

Proof: 作业

Theorem 1.3.2. Let P be the transition probability matrix for a connected Markov Chain, there is a unique probability vector π satisfying $\pi P = \pi$. Moreover, for any starting distribution, $\lim_{t\to\infty} a_t$ exists and equals π.

Proof. Consider the difference between a_t and $a_{t+1} = a_t P$:

$$\begin{aligned} a_t P - a_t &= \tfrac{1}{t}[p_0 P + p_1 P + \cdots + p_{t-1} P] - \tfrac{1}{t}[p_0 + p_1 + \cdots + p_{t-1}] \\ &= \tfrac{1}{t}[p_1 + p_2 + \cdots + p_t] - \tfrac{1}{t}[p_0 + p_1 + \cdots + p_{t-1}] \\ &= \tfrac{1}{t}(p_t - p_0) \end{aligned}$$

Thus, $b_t = a_t P - a_t$ satisfies $|b_t| \le \frac{2}{t}$ and goes to 0 as $t \to \infty$.

By Lemma 1.3.1, $A = [P - I, 1]$ has rank n.

Since the first n columns of A sum to 0, the $n \times n$ submatrix B of A consisting of all its columns except the first is invertible.

Let c_t be obtained from $b_t = a_t P - a_t$ by removing the first entry so that $a_t B = [c_t, 1]$.

Then $a_t = [c_t, 1]B^{-1}$.

Since $b_t \to 0, [c_t, 1] \to [0, 1]$ and $a_t \to [0, 1]B^{-1}$.

Since $a_t \to \pi$, we have that π is a probability vector.

Since $a_t[P - I] = b_t \to 0$, we get $\pi[P - I] = 0$.

Since A has rank n, this is the unique solution as required. □

Lemma 1.3.3. For a random walk on a strongly connected graph with probabilities on the edges, if the vector π satisfies $\pi_x p_{xy} = \pi_y p_{yx}$ for all x and y and $\sum_x \pi_x = 1$, then π is the stationary distribution of the walk.

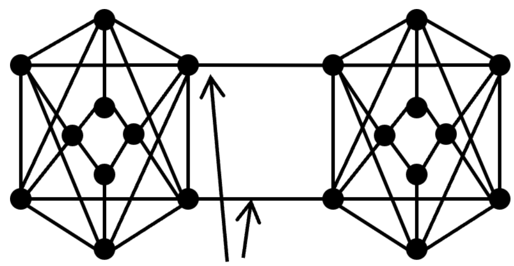
Proof. Since π satisfies $\pi_x p_{xy} = \pi_y p_{yx}$, summing both sides, $\pi_x = \sum_y \pi_y p_{yx}$ and hence π satisfies $\pi = \pi P$.

By Theorem 1.3.2 $\cdots\cdots$ $\square$

3 Convergence of Random Walks on Undirected Graphs

Next question: how fast the walk starts to reflect the stationary probability of the Markov process?

Example: this takes a long time to converge. The walk is unlikely to reach the narrow passage between the two halves.



We define below a combinatorial measure of constriction for a Markov Chain, called the normalized conductance.

Def 1.3.1. Fix $\varepsilon > 0$. The ε-mixing time of a Markov Chain is the minimum integer t such that for any starting distribution P_0, the 1-norm distance between the t-step running average probability distribution and the stationary distribution is at most ε.

$$|a_t - \pi| \leq \varepsilon$$

Def 1.3.2. For a subset S of vertices, let $\pi(S)$ denote $\sum_{x\in S} \pi_x$. The normalized conductance

$$\Phi(S) = \frac{\sum_{(x,y)\in(S,\overline{S})} \pi_x p_{xy}}{min(\pi(S), \pi(\overline{S}))}$$

$\overline{S} = V - S$. $\pi(S)$ is the probability that, at the stationary distribution, the Markov Chain will be at some state in S.

Def 1.3.3. The normalized conductance of the Markov Chain, denoted Φ, is defined by

$$\Phi = min_S \Phi(S)$$

Theorem 1.3.4. The ε-mixing time of a random walk on an undirected graph is

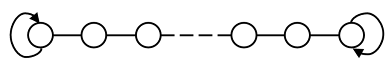
$$\Phi(\frac{\ln{(1/\pi_{min})}}{\Phi^2\varepsilon^3})$$

where π_{min} is the minimum stationary probability of any state.

Using Nomalized Conductance to prove Convergence.

Next, we apply theorem 1.3.4 to some examples to illustrate how the normalized conductance bounds the rate of convergence.

① A 1-dimensional lattice



n-vertex path with self-loops at the both ends.

The stationary probability is a uniform $\frac{1}{n}$ over all vertices.

The set with minimum normalized conductance is

–the set with probability $\pi \leq \frac{1}{2}$.

–the set consists of the first $\frac{n}{2}$ vertices.

The total conductance of edges from S to $\overline{S}$ is

$\pi_m p_{m,m+1} = \Omega(\frac{1}{n}), (m = \frac{n}{2})$

$\pi(S) = \frac{1}{2}$

Thus, $\Phi(S) = 2\pi_m p_{m,m+1} = \Omega(\frac{1}{n})$

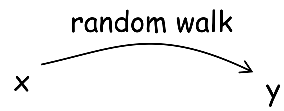
By theorem 1.3.4 for $\varepsilon = \frac{1}{100}$ after $O(n^2 \log n)$ steps, $||a_t - \pi||, \leq \frac{1}{100}$.

This graph does not have rapid covergence.

4 Random walks on Undirected Graphs with Unit Edge Weights

We use this special type of graph to answer:

- what is the expected time for a random walk starting at x to reach y.
- what is the expected time until the walk starting at x and return to x.
- what is the expected time to reach every vertex.

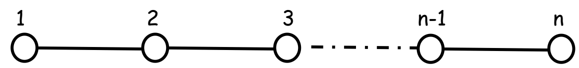


① **Hitting time**

h_{xy}-also called discovery time.

Lemma 1.3.5. The expected time for a random walk starting at one end of a path of n vertices to reach the other end is $\Theta H(n^2)$

Proof:



$h_{12} = 1$

$h_{i,i+1} = \frac{1}{2} + \frac{1}{2}(1 + h_{i-1,i+1})$

$= 1 + \frac{1}{2}(h_{i-1,i} + h_{i,i+1})$

$= 2 + hi - 1, i$

$\therefore h_{i,i+1} = 2i - 1, 2 \leq i \leq n - 1$

to get from 1 to n.

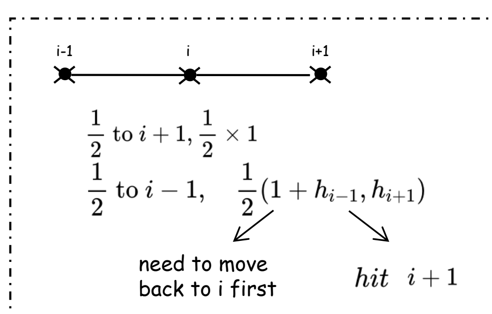
$h_{1,n} = \sum_{i=1}^{n-1} h_{i,i+1}$

$= \sum_{i=1}^{n-1}(2i - 1)$

$= 2\sum_{i=1}^{n-1} i - \sum_{i=1}^{n-1} 1$

$= 2\frac{n(n-1)}{2} - (n - 1)$

$= (n - 1)^2$



Lemma 1.3.6 Consider a random walk from vertex 1 to vertex n in a chain of n vertices. Let $t(i)$ be the expected time spent at vertex. Then

$$t(i) = \begin{cases} n-1, i=1 \\ 2(n-i), 2 \le i \le n-1 \\ 1, i=n \end{cases}$$

Proof Now $t(n) = 1$ since the walk stops when it reaches n. Half of the time when the walk is at $n-1$, it goes to n. Thus $t(n-1) = 2$. For $3 \le i \le n-1$

$$\begin{aligned} &t(i) = \tfrac{1}{2}[t(i-1) + t(i+1)] \\ &t(1) = \tfrac{1}{2}t(2) + 1 \\ &t(2) = t(1) + \tfrac{1}{2}t(3) \end{aligned}$$

Then we get

$t(i+1) = 2t(i) - t(i-1)$

$\therefore t(i) = 2(n-i), for 3 \le i \le n-1$

$t(2) = 2(n-2), t(1) = n-1.$

$\therefore$ The total time spent at vertices is

$n-1+2(1+2+...+n-2)+1 = (n-1+2\frac{(n-1)(n-2)}{2}+1) = (n-1)^2+1$

which is one more that h_{1n}.$\square$

② **Commute time**

$$commute(x, y) = h_{xy} + h_{yx}$$

③ **Cover time**

$Cover(x, G) \rightarrow$ expected time of a random walk starting at x to reach each vertex at least once.

$Cover(G) = \max_x cover(X, G)$

Theorem 1.3.7. Giver G with n vertices and m edges. Cover(G) is bounded above by $4m(n-1)$

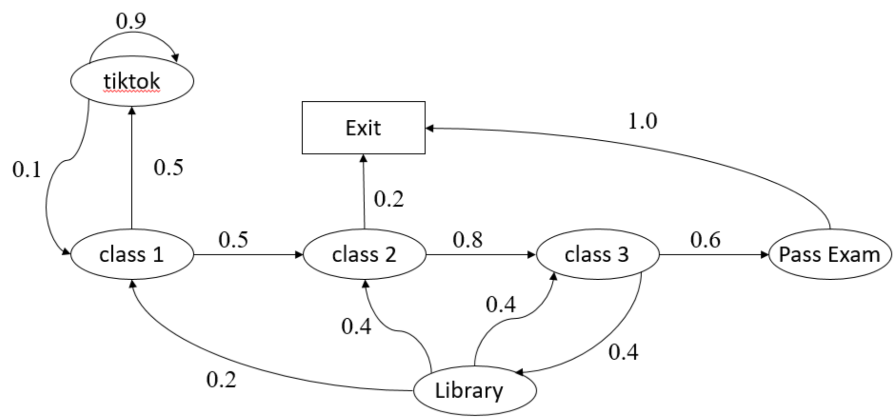
Proof. A depth first search starting from some vertex Z. T is the resulting depth first search spanning tree. The depth first search covers every vertex. Note that each edge in T is traversed twice, once in each direction.

$$Cover(Z, G) \le \sum_{(x,y)\in T,(y,x)\in T} h_{xy}$$

Corollary. If x and y are adjancent, $h_x y$ + $h_y x \leq 2m$, m is the number of edges.
The corollary implies $h_{xy} \leq 2m$.
Since there are $n-1$ edges in the dfs tree and each edge is traversed twice. Cover(Z) $\leq 4m(n-1)$
Thus, cover(G) $\leq 4m(n-1)$.

5 More about Markov

$\triangle$ **A simple Marker chain** $\langle S, P \rangle$ S **: State,** P **: Probability**,



A large number of paths.

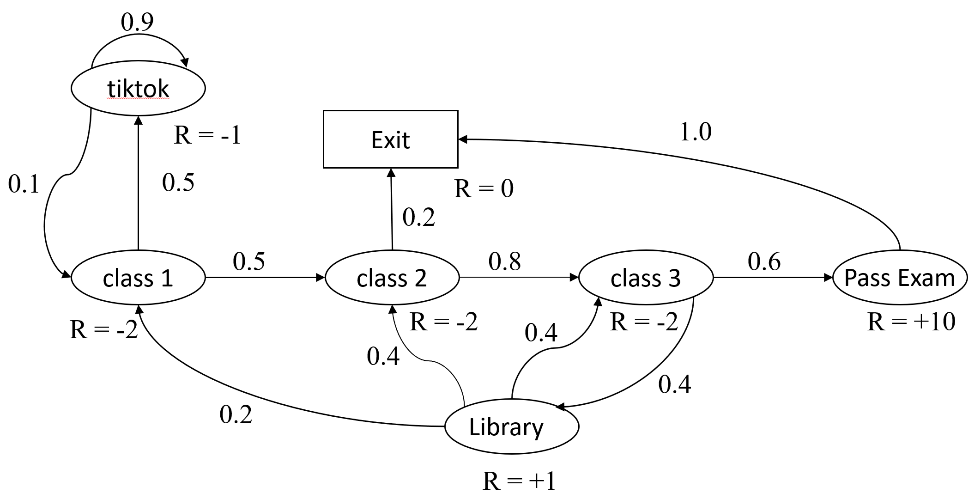
$\rightarrow$ $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3,$ pass

$\rightarrow$ C_1, Tiktok, $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3$, pass

$\rightarrow$ $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3,$ Library, C_2, pass

......

$\triangle$ **Markov Reward Process** $\langle S, P, R, \gamma \rangle$ R **: Reward,** γ**: Discount Factor**,



Total reward

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots$$

$$= \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}.$$

Value function of a state $V(s) = \mathbb{E}[G_t \mid S_t = s]$

$=$ The expectation of the reward from this state. i.e. average reward of different paths.

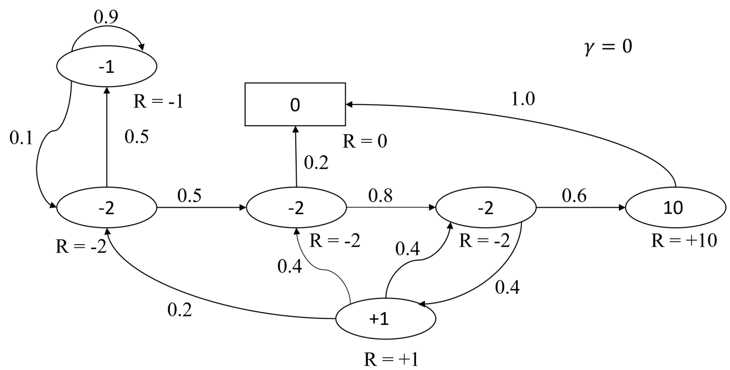
$C_1, C_2, C_3,$ Pass, Exit.

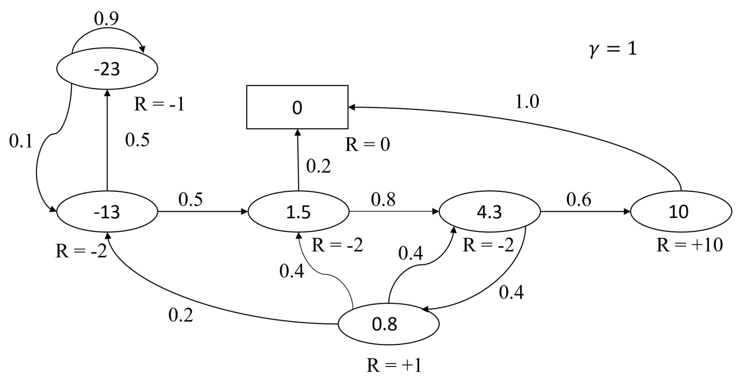
$S_1 = C_1$ with $\gamma = 1/2$

$$V_{C1} = -2 - 2 \cdot \frac{1}{2} - 2 \cdot \frac{1}{4} + 10 \cdot \frac{1}{8} = -2.25$$

$C_1,$ tiktok, tiktok,$C_1, C_2,$ Exit

$$V_{C_1} = -2 - 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{4} - 2 \cdot \frac{1}{8} - 2 \cdot \frac{1}{16} = -3.125$$





Bellman expectation equation.

$$
\begin{aligned}
V(s) &= \mathbb{E}\left[G_t \mid s_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \mid s_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma\left(R_{t+2} + \gamma R_{t+3} \cdots\right) \mid s_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma G_{t+1} \mid s_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma v\left(s_{t+1}\right)_{\mid} s_t = s\right]
\end{aligned}
$$

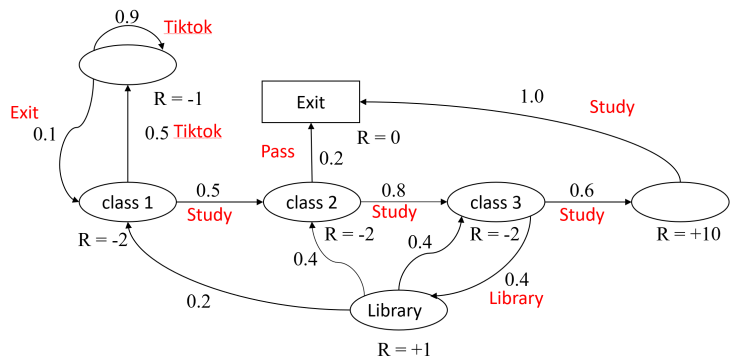
Use s' to denote the possible state of $t+1$,

$$V(s) = R_s + \gamma \sum_{s' \in S} P_{ss'} v\left(s'\right)$$

For class 3.

$$4.3 = -2 + 0.6 \times 10 + 0.4 \times 0.8$$

$\triangle$ **Markov Decision Process** $\langle S, A, P, R, \gamma\rangle$ A **: Action,**



Policy: probability distribution of taking actions.

$$\pi(a \mid s)=\mathbb{P}\left[A_t=a \mid st=s\right]$$

Given a MDP $M=\langle S, A, P, R, \gamma\rangle$ and a policy π. sequence $S_1, S_2, \cdots$ is a Markov process $\langle S, p^\pi\rangle$

Sate and reward sequence $S_1, R_2, S_2, \cdots$ is a Markov process $\langle S, P^\pi, R^\pi, \gamma\rangle$

Under policy π, probability of $s \rightarrow s'$ is

$$P_{ss^1}^\pi=\sum_{a \in A} \pi(a \mid s) P_{ss'}^a$$

Under policy π, the reward of s is

$$R_s^\pi=\sum_{a \in A} \pi(a \mid s) R_s^a$$

The value function:

$$V_\pi(s)=\mathbb{E}_\pi\left[G_t \mid s_t=s\right]$$

The policy-value function

$$q_\pi(s, a)=\mathbb{E}_\pi\left[G_t \mid s_t=s, A_t=a\right]$$

MDP Bellman function:

$$
\begin{aligned}
V_\pi(s) &= \mathbb{E}_\pi\left[R_{t+1}+\gamma v_\pi\left(s_{t+1}\right) \mid s_t=s\right] \\
q_\pi(s, a) &= \mathbb{E}_\pi\left[R_{t+1}+\gamma q_\pi\left(s_{t+1}, A_{t+1}\right) \mid s_t=s, A_t=a\right] \\
\therefore \quad V_\pi(s) &= \sum_{a \in A} \pi(a \mid s) q_\pi(s, a)
\end{aligned}
$$

And $q_\pi(s)=R_s^q+\gamma \sum_{s' \in S} P_{ss'}^a v_\pi\left(s'\right)$

$$
\therefore q_\pi(s)=R_s^a+\gamma \sum_{s' \in s} P_{ss'}^a \sum_{a' \in A} \pi\left(a' \mid s'\right) q \pi\left(s', a'\right)
$$

Optimal value function

$$
V_*(s)=\max_\pi V_\pi(s)
$$

Optimal action-value function

$$
q_*(s, a)=\max_\pi q_\pi(s \cdot a)
$$

We can use maximizing $q_*(s, a)$ to get the optimal policy π.

$$
\pi_*(a \mid s)= \begin{cases}1 & \text { if } a=\underset{a \in A}{\operatorname{argmax}} q_*(s, a) \\ 0 & \text { otherwise }\end{cases}
$$