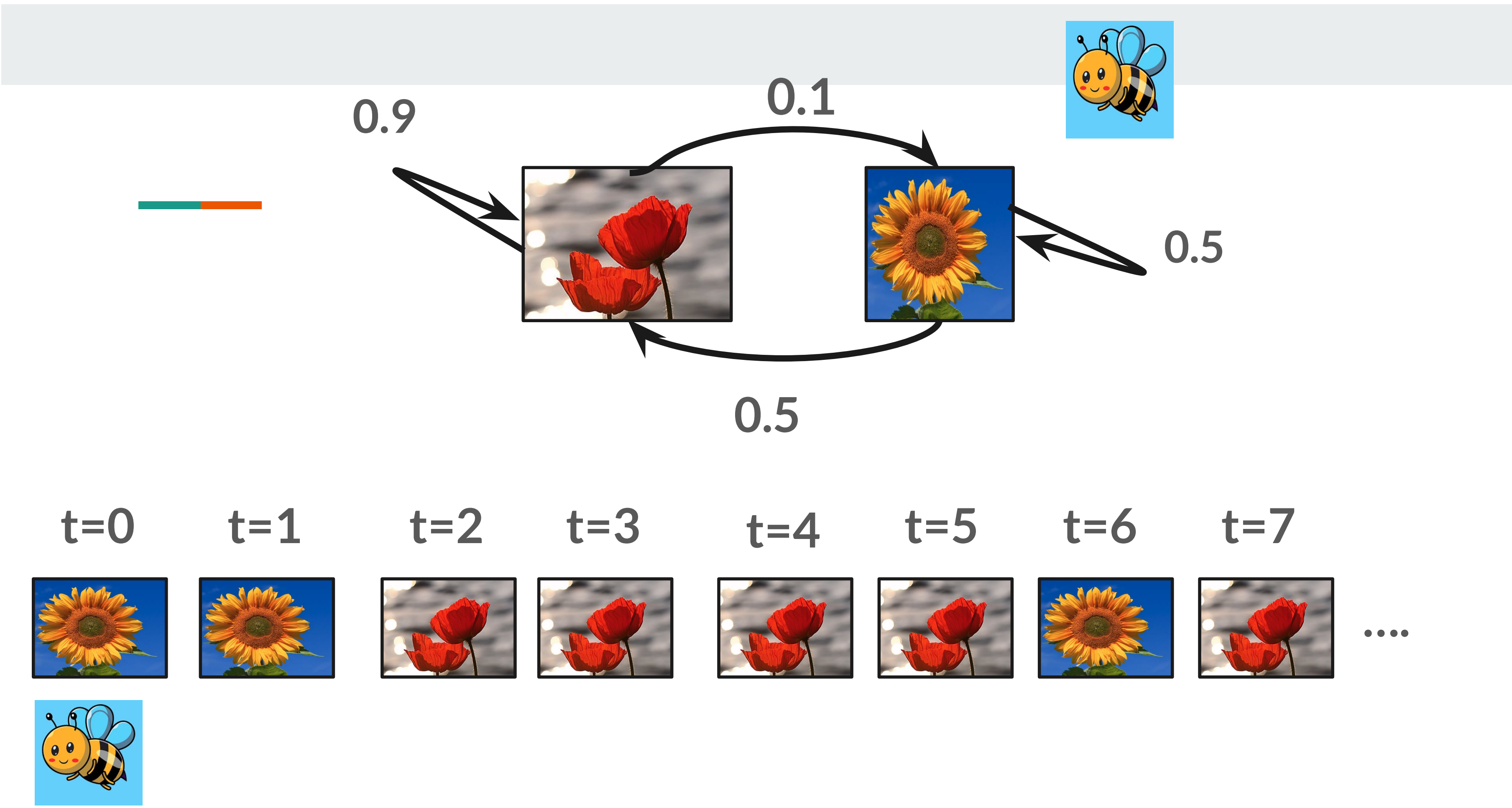
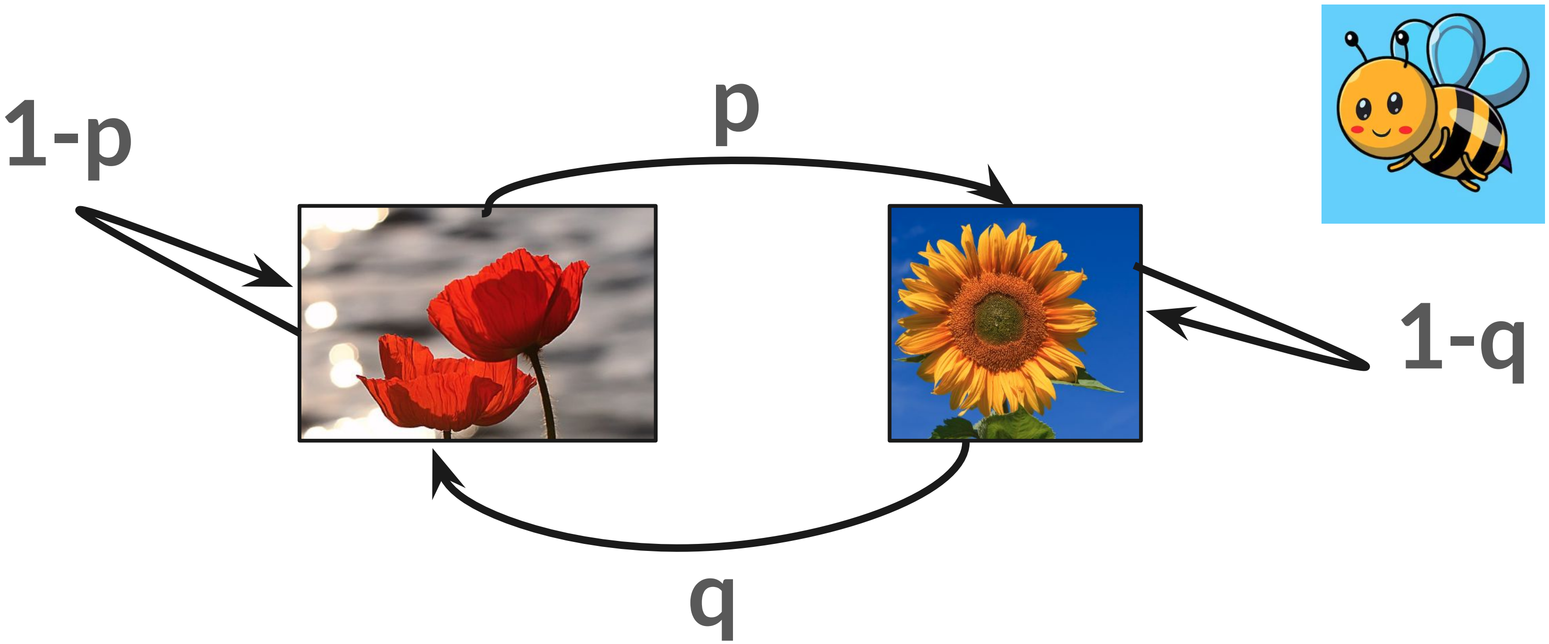
CS365
Foundations of Data Science
Markov Chains

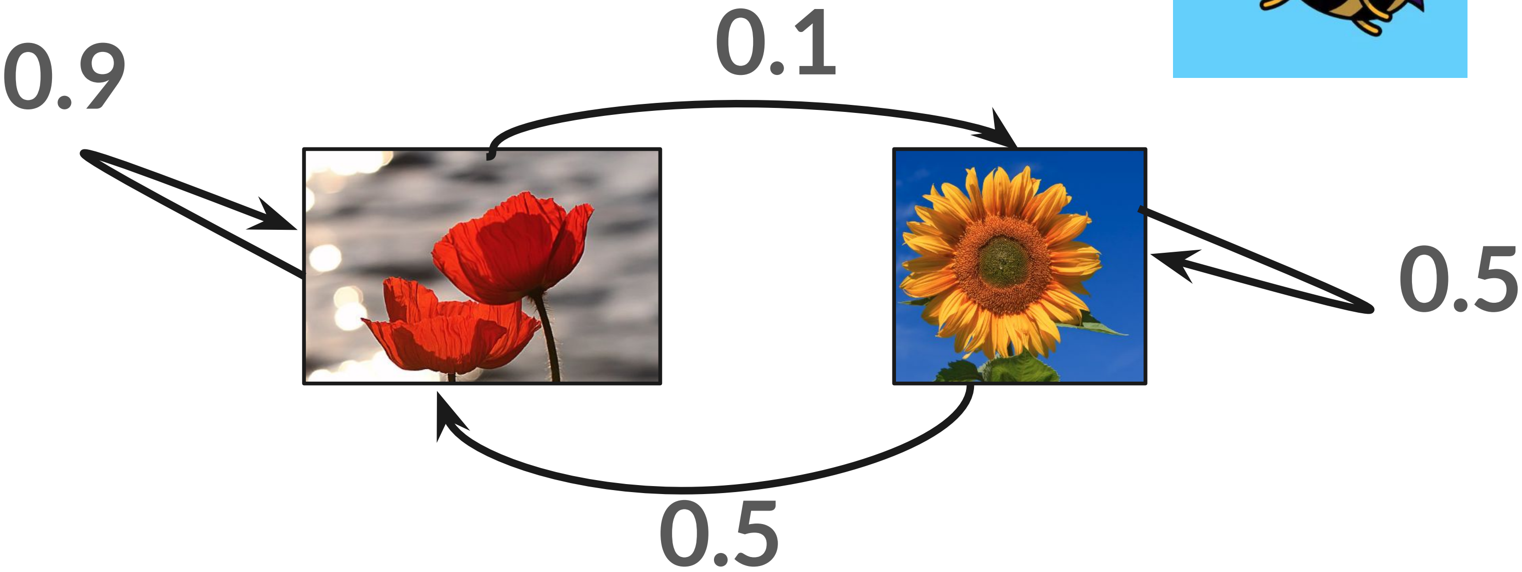
Charalampos E. Tsourakakis
ctsourak@bu.edu


0.9
0.1
0.5
0.5
t=0
t=1
t=2
t=3
t=4
t=5
t=6
t=7
....

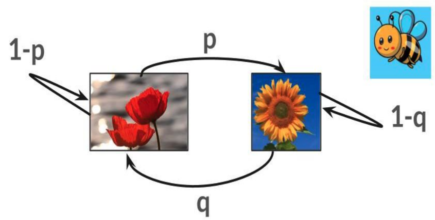
Markov chain



Markov chain



Transition matrix and starting probability distribution



$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

$\mu_0 = [Pr(X_0 =$  $), Pr(X_0 =$  $)]$

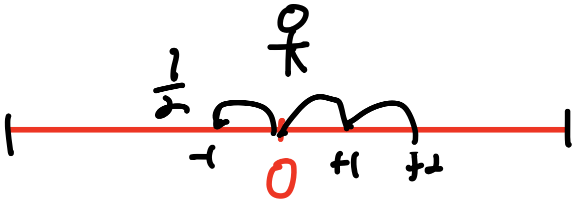
initial state

1 0

or

0 1

Drunkard's walk



Finite (Discrete Time) Markov Chain

state space is infinite

- Finite space Ω. Some examples:
  - In our toy example Ω={  ,  }

poppy

Definition

A sequence of random variables $(X_0, X_1, X_2, \ldots)$ is a Markov chain with state space Ω and transition matrix P if for all $x, y \in \Omega, t \geq 1$ and all events $H_{t-1} = \cap_{s=0}^{t-1}\{X_s = x_s\}$

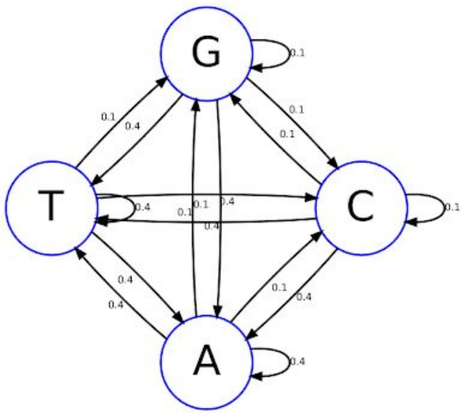
Satisfying $Pr(H_{t-1} \cap X_t = x) > 0$ we have

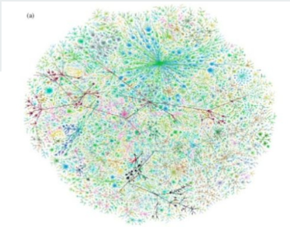
$$\boxed{Pr(X_{t+1} = y | H_{t-1} \cap \{X_t = x\}) = Pr(X_{t+1} = y | \{X_t = x\}) = P(x, y)}$$

History

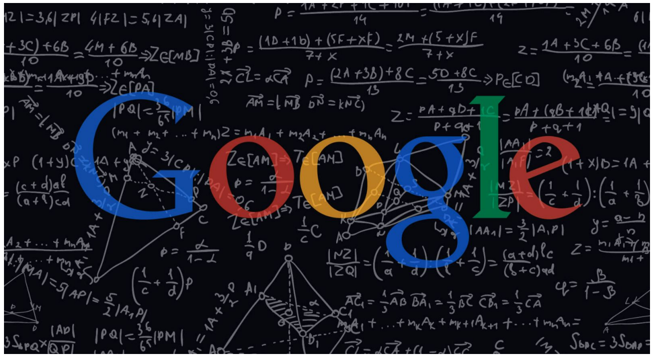
Finite Markov Chain

- Finite space Ω. Some examples:
  - In our toy example Ω={  ,  }
  - Ω = {set of all web pages}
  - Ω = {A, T, G, C}

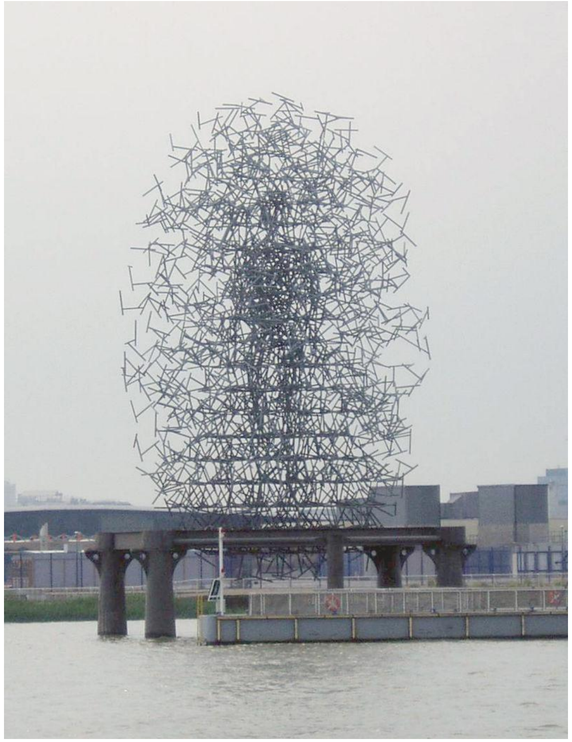




Web structure



Quantum Cloud



Antony Gormley

26 languages

Article | Talk | Read | Edit | View history | Tools

From Wikipedia, the free encyclopedia

Sir Antony Mark David Gormley OBE RA (born 30 August 1950) is a British sculptor.[1] His works include the *Angel of the North*, a public sculpture in Gateshead in the north of England, commissioned in 1994 and erected in February 1998; *Another Place* on Crosby Beach near Liverpool; and *Event Horizon*, a multipart site installation which premiered in London in 2007, then subsequently in Madison Square in New York City (2010), São Paulo, Brazil (2012), and Hong Kong (2015–16).

Sir
Antony Gormley
OBE RA

Gormley in 2011

Born	Antony Mark David Gormley 30 August 1950 (age 73) Hampstead, London, England
Education	Trinity College, Cambridge Saint Martin's School of Art Goldsmiths, University of London Slade School of Fine Art

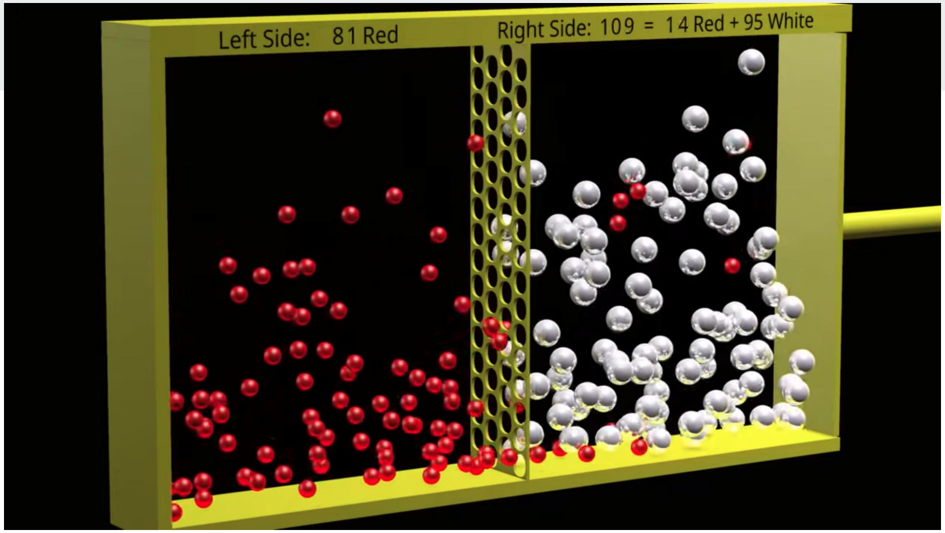
Early life [edit]

Gormley was born in Hampstead, London, the youngest of seven children, to a German mother (maiden name Brauninger) and a father of Irish descent.[2][3][4] His paternal grandfather was an Irish Catholic from Derry who settled in Walsall in Staffordshire.[5] The ancestral homeland of the Gormley Clan (Irish: *Ó Goirmleadhaigh*) in Ulster was east County Donegal and west County Tyrone,[6] with most people in both Derry and Strabane being of County Donegal origin. Gormley has stated that his parents chose his initials, "AMDG", to have the inference Ad maiorem Dei gloriam – "to the greater glory of God".[7]

Gormley grew up in a Roman Catholic[8] family living in Hampstead Garden Suburb. The family was wealthy, with a cook and a chauffeur, with a home overlooking the golf course; Gormley's father was an art lover.[4] He attended Ampleforth College, a Benedictine boarding school in Yorkshire,[4] before reading archaeology, anthropology, and the history of art at Trinity College, Cambridge, from 1968 to 1971.[4] He travelled to India and the Dominion of Ceylon / Sri Lanka to learn more about Buddhism between 1971 and 1974.[4]

After attending Saint Martin's School of Art and Goldsmiths in London from 1974, he completed his studies with a postgraduate course in sculpture at the Slade School of Fine Art, between 1977 and 1979. [*citation needed*]

Bernoulli model for osmosis



Two adjacent containers A and B each contain m particles; n are of type I and m are of type II. A particle is selected at random in each container. If they are of opposite types they are exchanged with probability α if the type I is in A, or with probability β if the type I is in B.

- State space: {0,1,2,...,n}
- Let X_n be the number of type I particles in A at time n.

Random walk on a graph

3x3 3x3 3x3 3x3

Adjacency matrix (A)

Starting point \ Destination	1	2	3
1	0	1	0
2	0	0	1
3	1	1	0

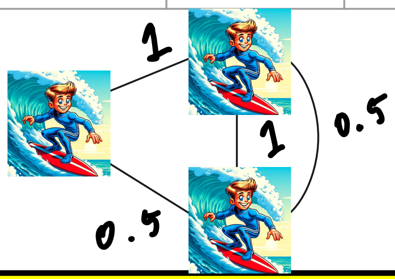
normalized

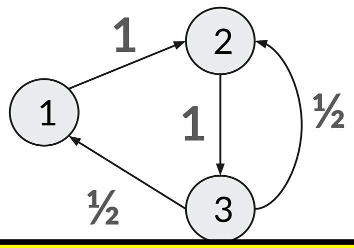
Transition matrix (P)

0	1	0
0	0	1
1/2	1/2	0

Hint: degree matrix D

1	0	0
0	1	0
0	0	2





How can we obtain P from A in an *algebraic way?* (or vice versa)

$$P = \begin{bmatrix} d_1 & & \\ & d_2 & \\ & & d_3 \end{bmatrix} A$$ inverse

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} A$$

$P = D^{-1}A$

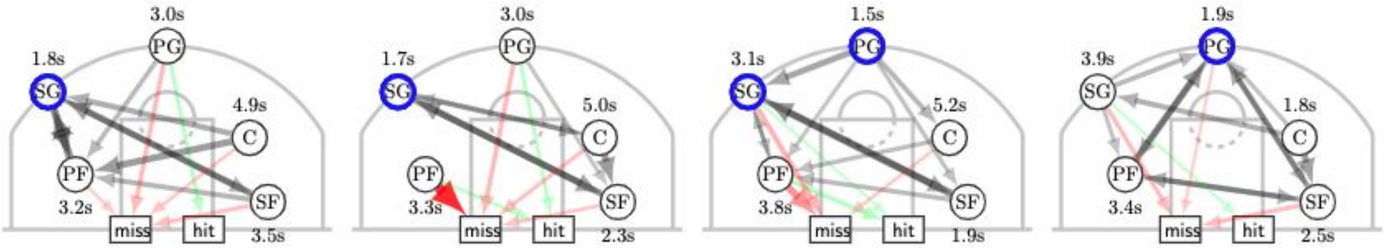
$A = DP$

$$\begin{bmatrix} 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix}$$

$$\begin{bmatrix} d_1\alpha_{11} & d_1\alpha_{12} & d_1\alpha_{13} \\ d_2\alpha_{21} & d_2\alpha_{22} & d_2\alpha_{23} \\ d_3\alpha_{31} & d_3\alpha_{32} & d_3\alpha_{33} \end{bmatrix}$$

Markovletics

- Mixtures of Markov chains



- Learning offensive strategies of teams in NBA from passing game.

Spaeh and T. (Webconf 2024)

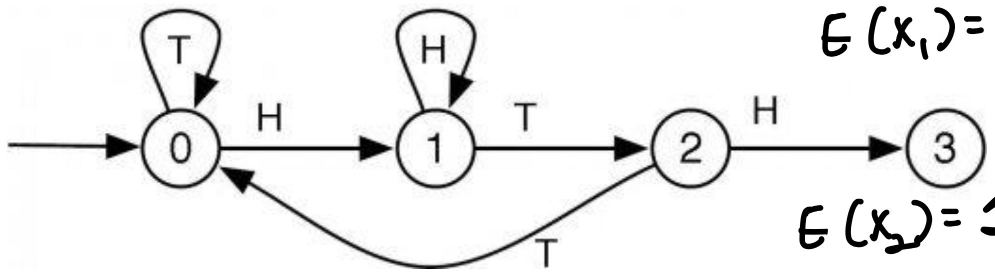
Waiting for HEADS-TAILS-HEADS (HTH) in a row

- Suppose we toss a coin consecutively until we observe the sequence of HEADS-TAILS-HEADS.

- Intuition: is the expected number of tosses for HTH equal to the expected number of tosses to observe HHH?

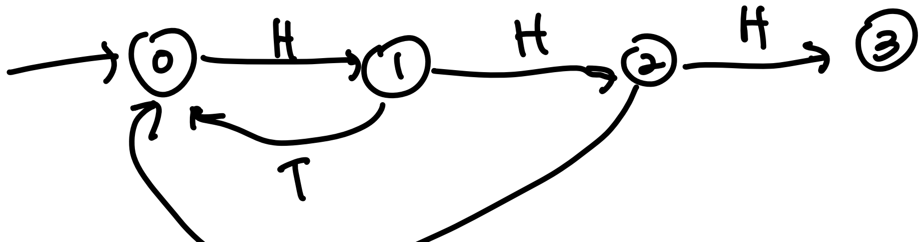
$E(X_0) = 1 + \frac{1}{2}(E(X_0) + E(X_1))$

$E(X_1) = 1 + \frac{1}{2}(E(X_1) + E(X_2))$



$E(X_2) = 1 + \frac{1}{2}(E(X_0) + E(X_3))$

$E(X_0) = 0$



Comprehension questions

- Suppose on Monday, the bee is in the poppy (i.e., $X_0 = 0$)
  - $\Pr(X_1 = 0 \mid X_0 = 0) = ?$ $1-p$
  - $\Pr(X_1 = 1 \mid X_0 = 0) = ?$ p
  - $\Pr(X_2 = 0 \mid X_0 = 0) = ?$
  - More generally, what is the probability $\Pr(X_t = 0 \mid X_0 = 0)$?

$$\begin{array}{c|cc} \text{Start} \backslash \text{Dest} & 0 & 1 \\ \hline 0 & 1-p & p \\ 1 & q & 1-q \end{array}$$

$$p \times \begin{cases} 0 \to \overset{2}{0} \to \overset{0}{0} & (1-p)^2 \\ & + \\ 0 \to 1 \to 0 & p \times q \end{cases}$$

Probability Distribution after t steps

- Let μ_t be the probability distribution of the Markov chain after t steps.

- Theorem: $\mu_t = \mu_0 P^t$

Proof on whiteboard.

(tomorrow) = (today) * transition matrix P

$$\mu_{t+1} = \begin{bmatrix} Pr(X_{t+1} = 1) \\ Pr(X_{t+1} = 2) \\ \\ Pr(X_{t+n} = n) \end{bmatrix} \qquad \mu_{t+1} = \mu_t P, \quad \mu_{t+1} = P^T \mu_t$$

$$\mu_{t+1} = \left[Pr(X_{t+1}=0),\ Pr(X_{t+1}=1) \right]$$

$$Pr(X_{t+1}=0) = Pr(X_{t+1}=0 \mid X_t=0)\,Pr(X_t=0) \quad = (1-p)\,Pr(X_t=0)$$

$$Pr(X_{t+1}=0 \mid X_t=1)\,Pr(X_t=1) \quad = (q)\,Pr(X_t=1)$$

Stationary distribution

- From the proof, we saw that $\mu_{t+1}=\mu_t P$.

- What happens when t->inf?

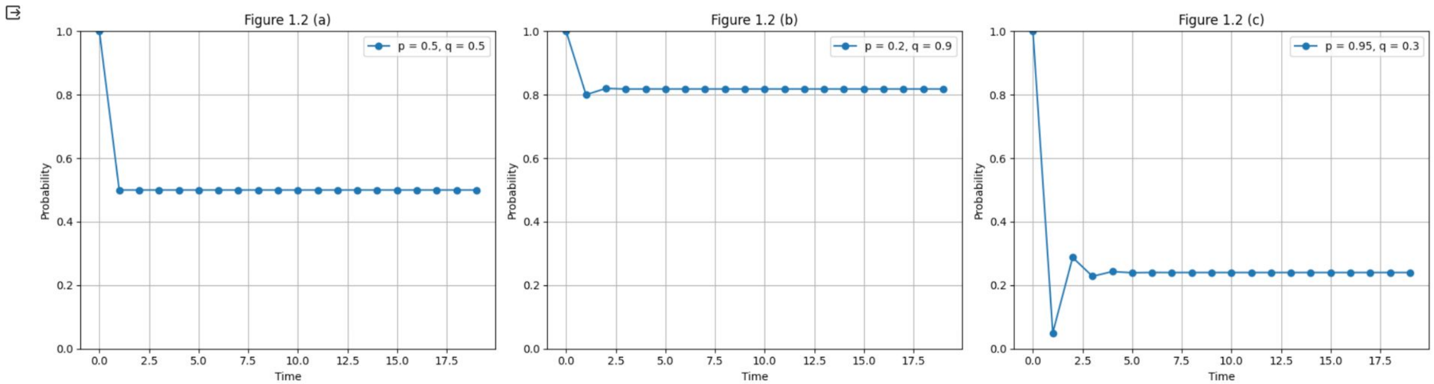
  Define for all t>=0 $\Delta_t = \mu_t(0) - \dfrac{q}{p+q}$

  We obtain:

$$\Delta_{t+1} = \mu_t(0)(1-p) + (1-\mu_t(0))(q) - \frac{q}{p+q} = (1-p-q)\Delta_t.$$

Stationary Distribution

Colab notebook [here]



Stationary distribution

$$\Delta_{t+1} = \mu_t(0)(1-p) + (1-\mu_t(0))(q) - \frac{q}{p+q} = (1-p-q)\Delta_t.$$

Notice that independently of Δ_0

$$\Delta_t = (1-p-q)^t \Delta_0$$

We conclude that when 0<p,q<1 then $\lim_{t \to +\infty} \mu_t(0) = \frac{q}{p+q}$

Similarly $\lim_{t \to +\infty} \mu_t(1) = \frac{p}{p+q}$

$$\lim_{t \to \infty} \mu_t = \mu$$

$$\mu = \mu P$$

$$(\mu_0, \mu_1) = (\mu_0, \mu_1) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

$\mu_0 = \mu_0(1-p) + q\mu_1$

$\mu_0 p = q\mu_1$

$(\mu_0 = 1$ then $\mu_1 = \frac{p}{q})$

because they have to be add up to 1 (μ_0, μ_1)

Stationary distribution

- Assuming the Markov chain in "well-behaving" for some large enough t, $\mu_t = \mu_{t+1}$

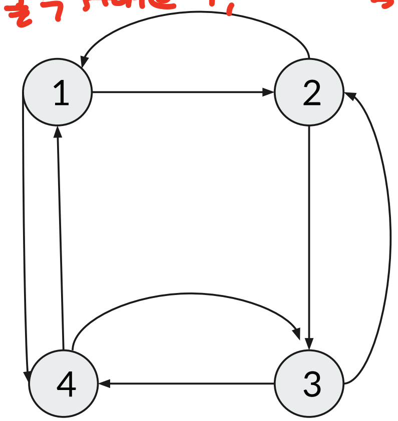
- Therefore, $\mu P = \mu$
  - μ is called the stationary distribution of the Markov chain.
  - Google's Pagerank intuition:
    The stationary distribution at a state/node corresponds to the proportion of time a random surfer spends at that state/node. Thus it is related to the importance of that state/node.

Questions

1. Can a random walk on the following graph have a stationary distribution?

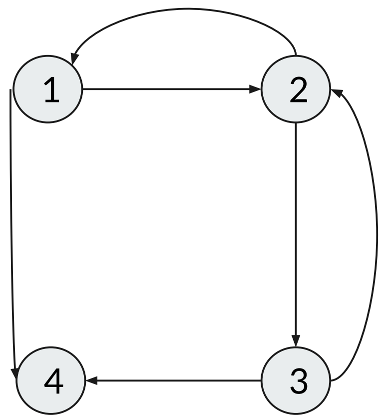
Nope pariotery grap?

홀수 state 1, 3 짝수 state 에 2, 4에 가능



Questions

2. Can a random walk on the following graph have a stationary distribution?



“Well-behaving” Markov chain

Can a random walk on the following graph have a stationary distribution?

1. **Irreducible**: There is a path from every node to every other node

2. **Aperiodic**: Define $T(x)=\{t>=1: P^t(x,x) > 0\}$, i.e., the set of times when it is possible for the chain to return to starting position x. The period of state x is the gcd of T(x). If all states have period 1, then the chain is aperiodic.

Lazy random walk

A common trick to avoid the periodicity problem is to create a lazy version of the random walk.

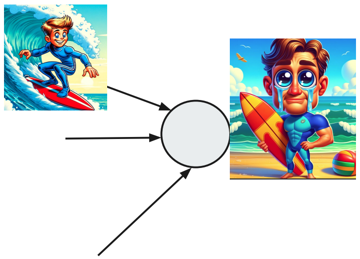
- Let P be the transition matrix.
- Define the lazy random walk as $P' = (P+I)/2$.
- In simple words the modified random walk defined by P':
  - The surfer stays at the current vertex with probability ½
  - The surfer takes a step of the original random walk with probability ½

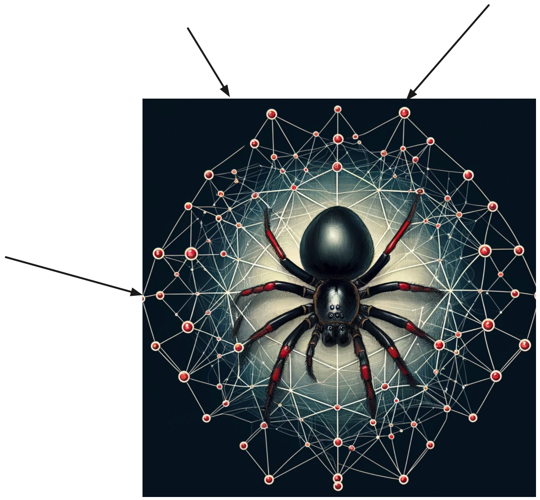
Pagerank

Is the web graph is irreducible?

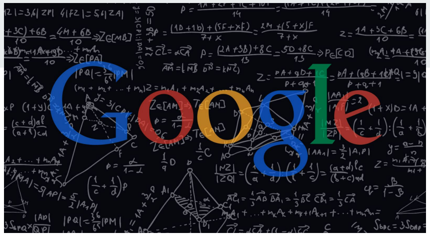
- Absolutely not!

Dead-ends and spider traps





Pagerank



Key idea:

- At any time-step the random surfer jumps (teleport) to any other node with probability c (e.g., c=0.05)

- Or jumps to its direct neighbors with total probability 1-c (e.g., 0.95).

$$\tilde{P} = (1 - c)P + c \cdot \frac{1}{n}\mathbf{1}\mathbf{1}^T$$

- The pagerank vector v is the solution to where r is the uniform distribution vector over all web page

$$\vec{pr} = (1 - c)\vec{pr}P + c\vec{r}$$