CS145: Probability & Computing

Lecture 18: Markov Chains, Recurrence, Stationary Distributions

Brown University Computer Science

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

CS145: Lecture 18 Outline

- Markov Chains over Multiple Time Steps
- Markov State Classifications and Durations
- Steady-State Behavior and Equilibrium Distributions
- Absorption in Markov Chains

Finite Markov Chains

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \quad \cdots$$

Finite State: $X_t \in \{1, \ldots, m\}$

Markov Property: *Given the current state, the past & future are independent.*

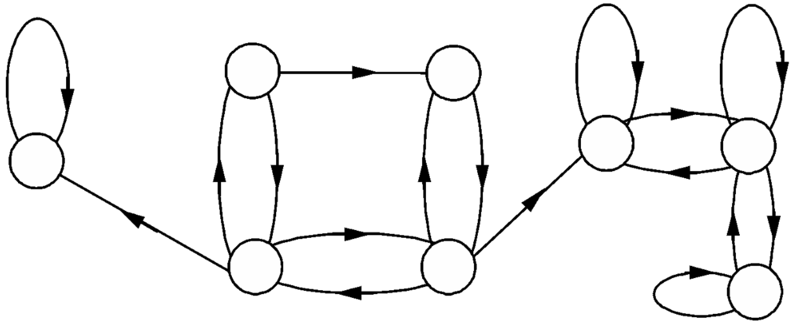
$$P(X_0, X_1, \ldots, X_n) = P(X_0) \prod_{t=1}^{n} P(X_t \mid X_{t-1})$$

State Transition Matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

State Transition Diagram:



Multi-Step State Transitions

$$\pi_{ti} = P(X_t = i) \qquad \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T$$

After *n* time steps:

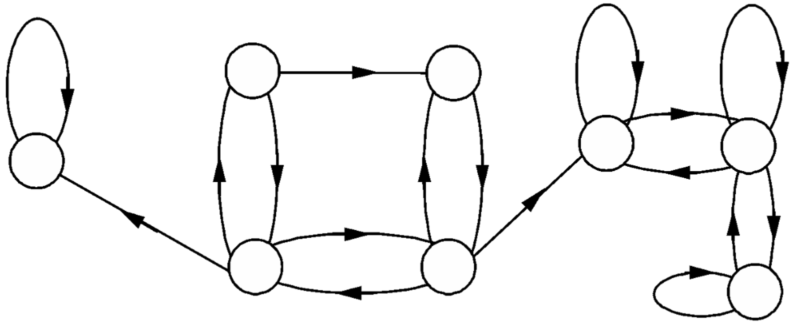
$$\pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n$$

State Transition Matrix:

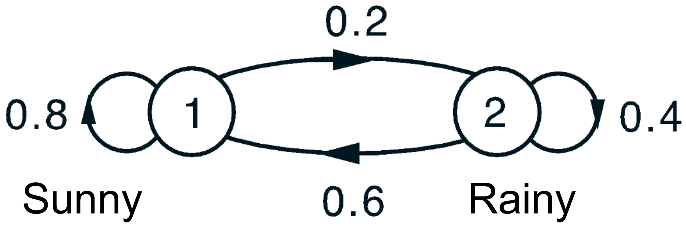
$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

State Transition Diagram:



Example: Sunny or Rainy?



$$r_{ij}(n) = P(X_n = j \mid X_0 = i) = P^n$$

$r_{ij}(1)$

	S	R
S	0.8	0.2
R	0.6	0.4

$r_{ij}(2)$

.76	.24
.72	.28

$r_{ij}(3)$

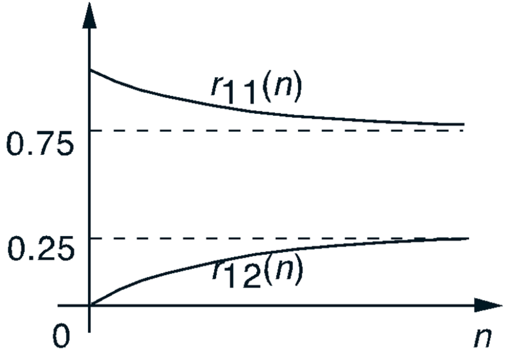
.752	.248
.744	.256

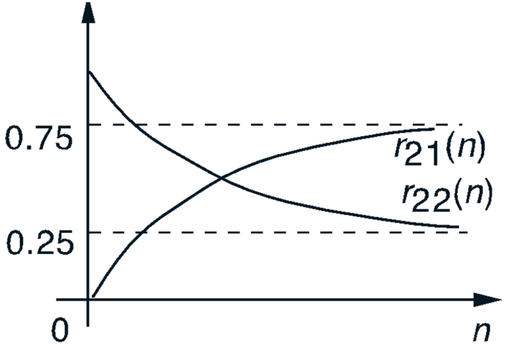
$r_{ij}(4)$

.7504	.2496
.7488	.2512

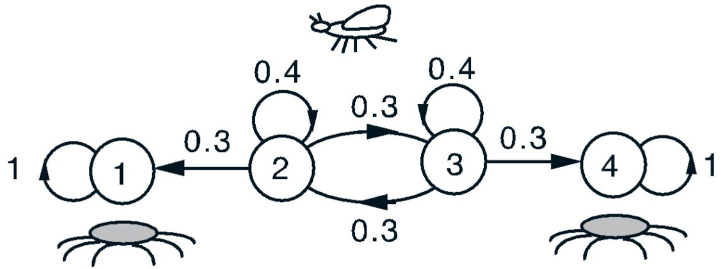
$r_{ij}(5)$

.7501	.2499
.7498	.2502





Example: Spiders and the Fly



	1	2	3	4
1	1.0	0	0	0
2	0.3	0.4	0.3	0
3	0	0.3	0.4	0.3
4	0	0	0	1.0

p_{ij}

$$r_{ij}(n) = P(X_n = j \mid X_0 = i) = P^n$$

	1	2	3	4
1	1.0	0	0	0
2	0.3	0.4	0.3	0
3	0	0.3	0.4	0.3
4	0	0	0	1.0

$r_{ij}(1)$

1.0	0	0	0
.42	.25	.24	.09
.09	.24	.25	.42
0	0	0	1.0

$r_{ij}(2)$

1.0	0	0	0
.50	.17	.17	.16
.16	.17	.17	.50
0	0	0	1.0

$r_{ij}(3)$

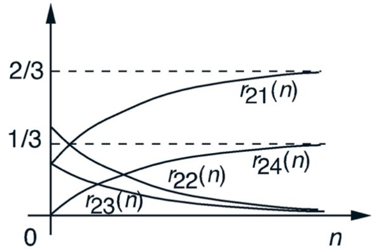
1.0	0	0	0
.55	.12	.12	.21
.21	.12	.12	.55
0	0	0	1.0

$r_{ij}(4)$

....

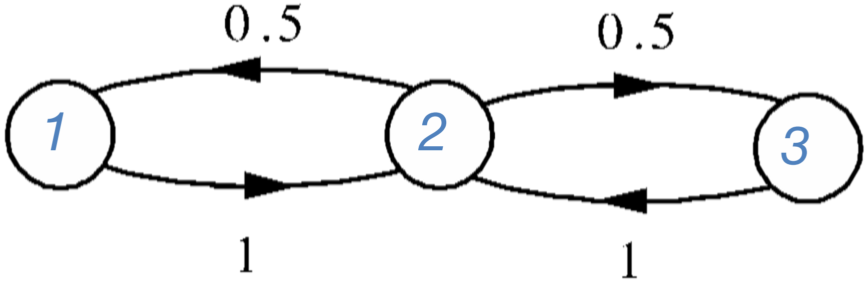
1.0	0	0	0
2/3	0	0	1/3
1/3	0	0	2/3
0	0	0	1.0

$r_{ij}(\infty)$



Markov Convergence Questions

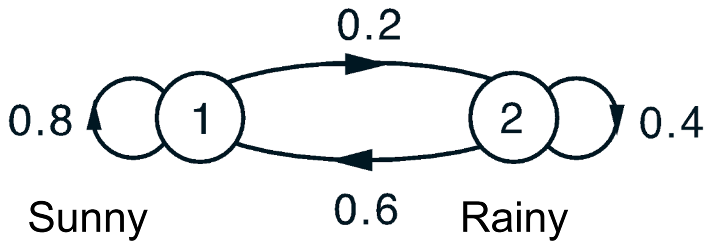
As $n \to \infty$, does $r_{ij}(n)$ converge to something?

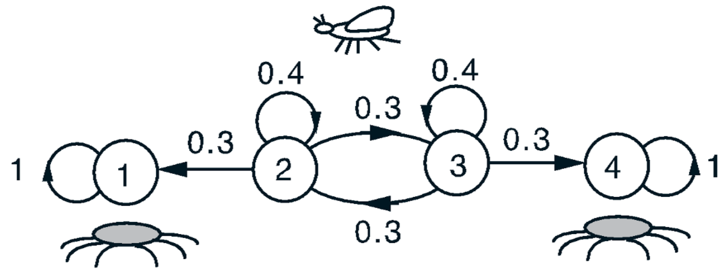


$r_{11}(n) =$ $r_{22}(n) =$ $r_{33}(n) =$

Markov Convergence Questions

As $n \to \infty$, does $r_{ij}(n)$ depend on the initial state i?





	S	R
S	0.8	0.2
R	0.6	0.4

$r_{ij}(1)$

....

0.75	0.25
0.75	0.25

$r_{ij}(\infty)$

	1	2	3	4
1	1.0	0	0	0
2	0.3	0.4	0.3	0
3	0	0.3	0.4	0.3
4	0	0	0	1.0

$r_{ij}(1)$

....

1.0	0	0	0
2/3	0	0	1/3
1/3	0	0	2/3
0	0	0	1.0

$r_{ij}(\infty)$

CS145: Lecture 18 Outline

- Markov Chains over Multiple Time Steps
- Markov State Classifications and Durations
- Steady-State Behavior and Equilibrium Distributions
- Absorption in Markov Chains

Classification of Markov Chain States

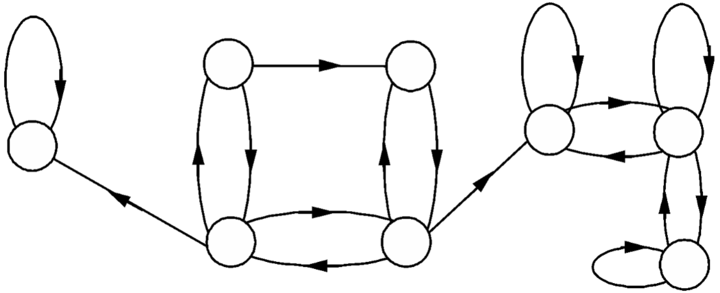
- **Accessible:** State *j* is accessible from state i if $r_{ij}(n)$ is positive for some *n*. The set of states accessible from state *i* then equals

$$A(i) = \{j \mid r_{ij}(n) > 0 \text{ for at least one time } n\}$$

- **Recurrent:** If a recurrent state is visited once, it will be visited an infinite number of times. State *i* is recurrent if and only if

$$\text{for all } j \in A(i), i \in A(j).$$

- **Transient:** Any state that is *not* recurrent is transient, and with probability one will only be visited a finite number of times



- **Absorbing:** A state *i* is absorbing if the process never leaves that state: $p_{ii} = 1$

Classification of Markov Chain States

- **Accessible:** State *j* is accessible from state i if $r_{ij}(n)$ is positive for some *n*. The set of states accessible from state *i* then equals

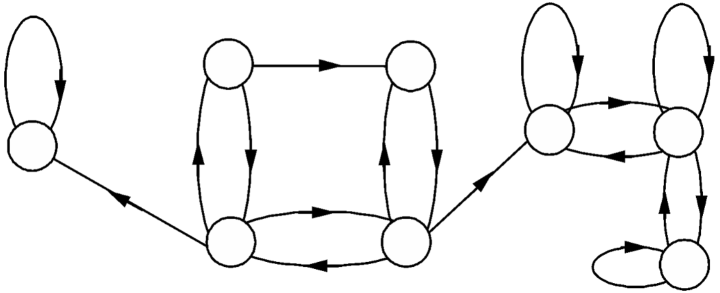
$$A(i) = \{j \mid r_{ij}(n) > 0 \text{ for at least one time } n\}$$

- **Recurrent:** If a recurrent state is visited once, it will be visited an infinite number of times. State *i* is recurrent if and only if

$$\text{for all } j \in A(i), i \in A(j).$$

- **Recurrent Class:** A collection of recurrent states for which all pairs are accessible from each other, or *communicate*

A pair of states (i,j) communicate if and only if the state transition diagram has directed paths from i to j, and from j to i.



Classification of Markov Chain States

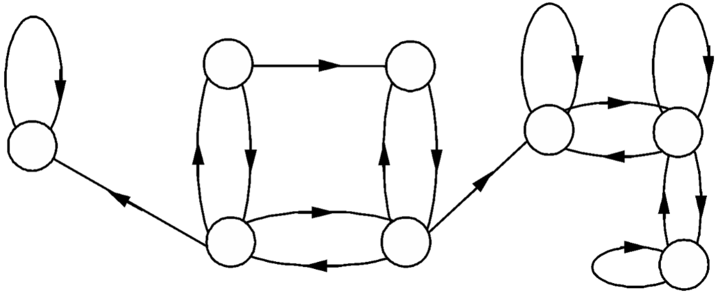
- **Accessible:** State *j* is accessible from state i if $r_{ij}(n)$ is positive for some *n*. The set of states accessible from state *i* then equals

$$A(i) = \{j \mid r_{ij}(n) > 0 \text{ for at least one time } n\}$$

- **Recurrent:** If a recurrent state is visited once, it will be visited an infinite number of times. State *i* is recurrent if and only if

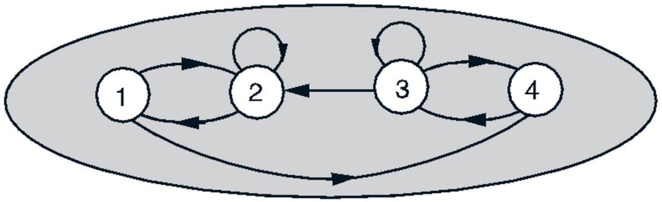
$$\text{for all } j \in A(i), i \in A(j).$$

- **Recurrent Class:** A collection of recurrent states for which all pairs are accessible from each other, or *communicate*



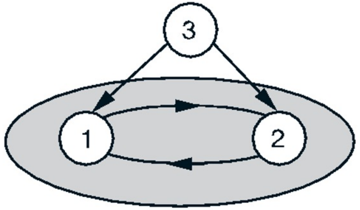
- **Irreducible:** A Markov chain is irreducible if it has only one recurrent class (plus possibly some transient states)

Examples of Markov Chain Decompositions



Irreducible

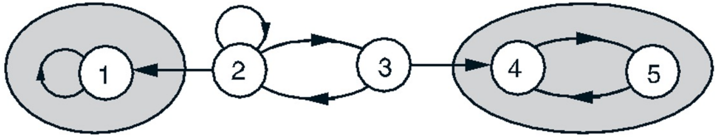
Single class of recurrent states



Irreducible

Single class of recurrent states (1 and 2)
and one transient state (3)

Absorbing State

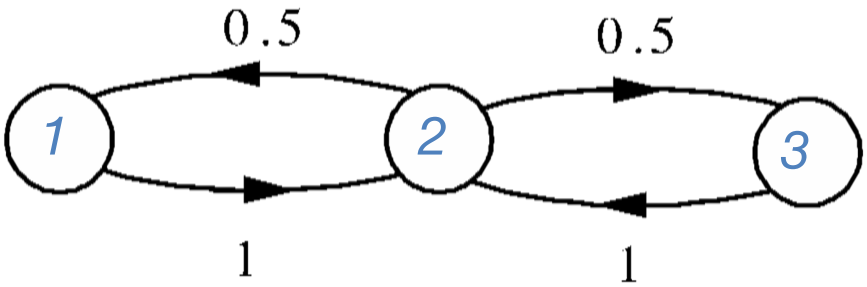


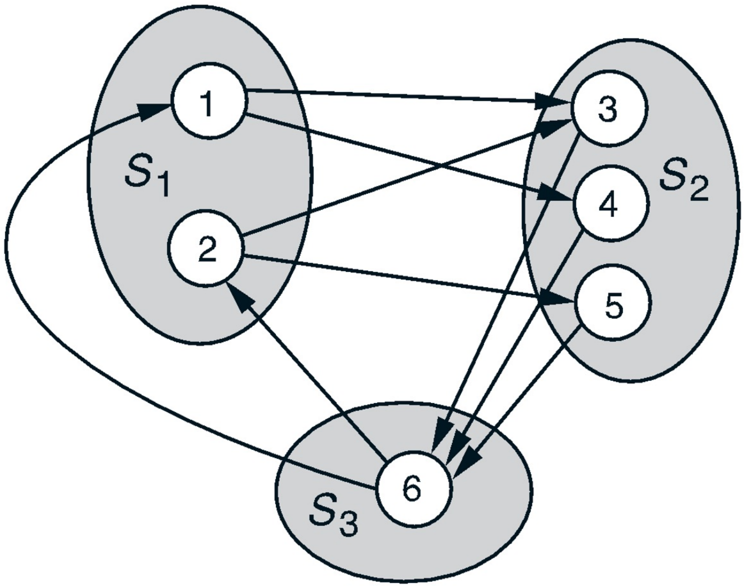
Not Irreducible (Reducible)

Two classes of recurrent states
(class of state1 and class of states 4 and 5)
and two transient states (2 and 3)

Periodic States

- **Periodic:** A state j in a Markov chain is periodic if there exists an integer $d > 1$ such $P(X_{t+s}=j \mid X_t=j) = 0$ unless s is divisible by d.
- **Aperiodic:** A Markov chain in which *no* recurrent states are periodic.





CS145: Lecture 18 Outline

- Markov Chains over Multiple Time Steps
- Markov State Classifications and Durations
- Steady-State Behavior and Equilibrium Distributions
- Absorption in Markov Chains

Multi-Step State Transitions

$$\pi_{ti} = P(X_t = i) \qquad \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T$$

After *n* time steps:

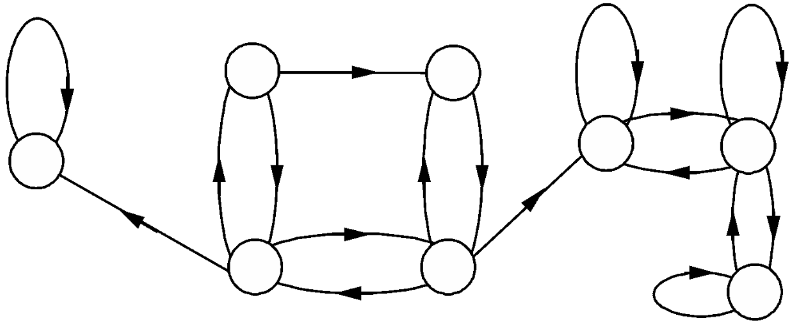
$$\pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T PP = \pi_0^T P^n$$

State Transition Matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

State Transition Diagram:



Steady-State (Equilibrium) Distribution

$$\pi_{ti} = P(X_t = i) \qquad \pi_t = [\pi_{t1}, \pi_{t2}, \ldots, \pi_{tm}]^T$$

After *n* time steps:

$$\pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T PP = \pi_0^T P^n$$

If state distribution converges: $||\pi_n - \pi_{n-1}|| \to 0$

The equilibrium distribution is the unique solution to m+1 linear equations:

$$\pi_j^* = \sum_{i=1}^{m} \pi_i^* p_{ij}, \qquad j = 1, \ldots, m. \qquad \sum_{i=1}^{m} \pi_i^* = 1.$$

$$\pi^* = P^T \pi^*$$

Alternative Interpretation: Eigenvector of state transition matrix whose eigenvalue is 1.

Conditions for Steady-State Distribution

A unique steady-state distribution exists if and only if:

- The Markov chain is *aperiodic* (not periodic).
- The Markov chain is *irreducible* (single recurrent class).

Properties of steady-state distributions:

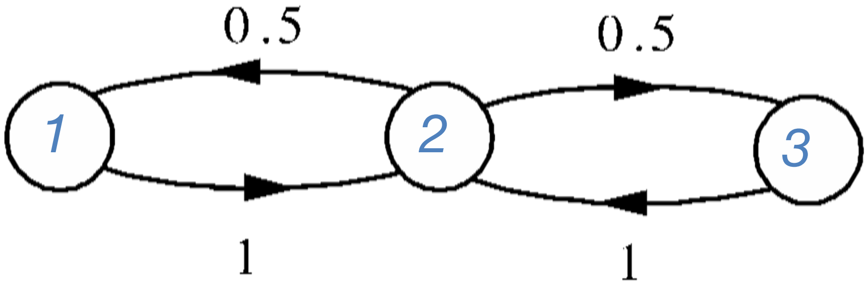
- For any transient state *i*, $\pi_i^* = 0$
- For all state pairs *(i,j)*, $\lim_{n\to\infty} r_{ij}(n) = \pi_j^*$

Linear equations defining unique steady-state distribution:

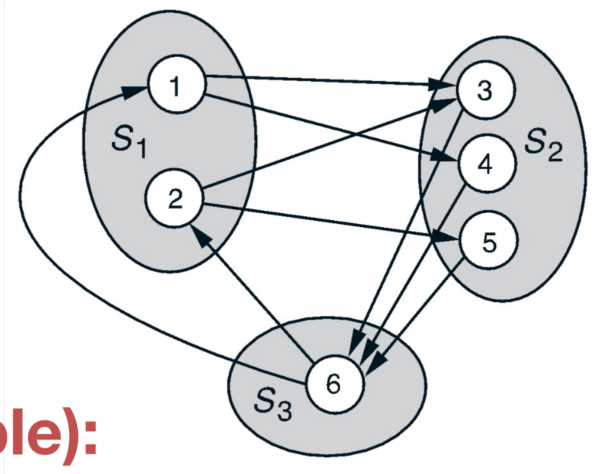
$$\pi_j^* = \sum_{i=1}^{m} \pi_i^* p_{ij}, \qquad j = 1, \ldots, m. \qquad\qquad \sum_{i=1}^{m} \pi_i^* = 1.$$

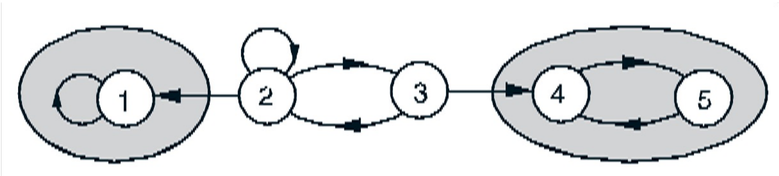
Violations of Steady-State Conditions

Periodic (not aperiodic) Markov chains:

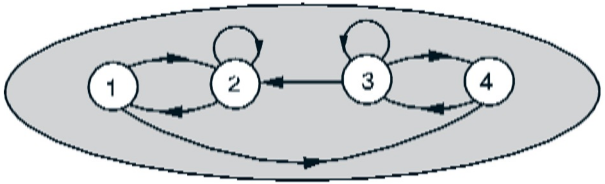


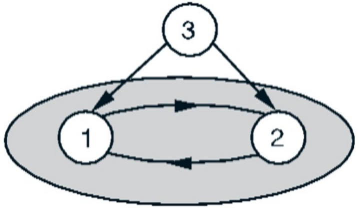
Multiple recurrent classes (not irreducible):





Both aperiodic and irreducible:

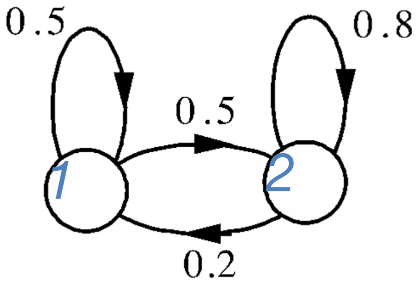




Transient state:

$$\pi_3^* = 0$$

Example: 2-State Stationary Distribution



$$(2/7, 5/7)\begin{pmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 2/7 \\ 5/7 \end{pmatrix}$$

$\pi_1 = 2/7,\ \pi_2 = 5/7$

- Assume process starts at state 1. $P(X_1 = 1) = 1$
- $\mathbf{P}(X_1 = 1,$ and $X_{100} = 1) =$ $\approx \pi_1 = \dfrac{2}{7}$
- $\mathbf{P}(X_{100} = 1$ and $X_{101} = 2)$ $\approx \pi_1 p_{12} = \dfrac{2}{7} \cdot \dfrac{1}{2} = \dfrac{1}{7}$

In Finite State Space Markov Chain

A finite state space Markov chain has a stationary distribution if and only if it is aperiodic, and its corresponding directed graph has one strongly connected component.

Example: Card Shuffling

- Markov chain states: All *n!* orderings of *n* distinct cards
- Transitions: Pick one of the *n* cards uniformly at random, and move that card to the top of the deck.
- *Stationary distribution:* Uniform distribution over permutations

Proof.

Consider a state $C = c_1, \ldots, c_n$. The chain can move to this state from the n states $C(j) = c_2, \ldots, c_1, \ldots, c_n$, where card c_1 is in place j of the deck. There are n such states, and each has probability $1/n$ to move to state C.

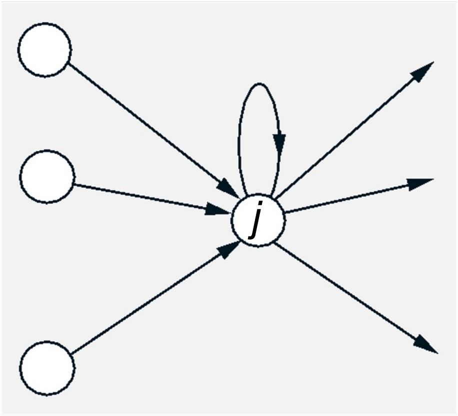
$$\pi_C = \sum_{j=1}^{n} \pi_{C(j)} \frac{1}{n}$$

$$\frac{1}{n!} = \sum_{j=1}^{n} \frac{1}{n} \frac{1}{n!}$$

Visit Frequency Interpretation

$$\pi_j = \sum_k \pi_k p_{kj}$$

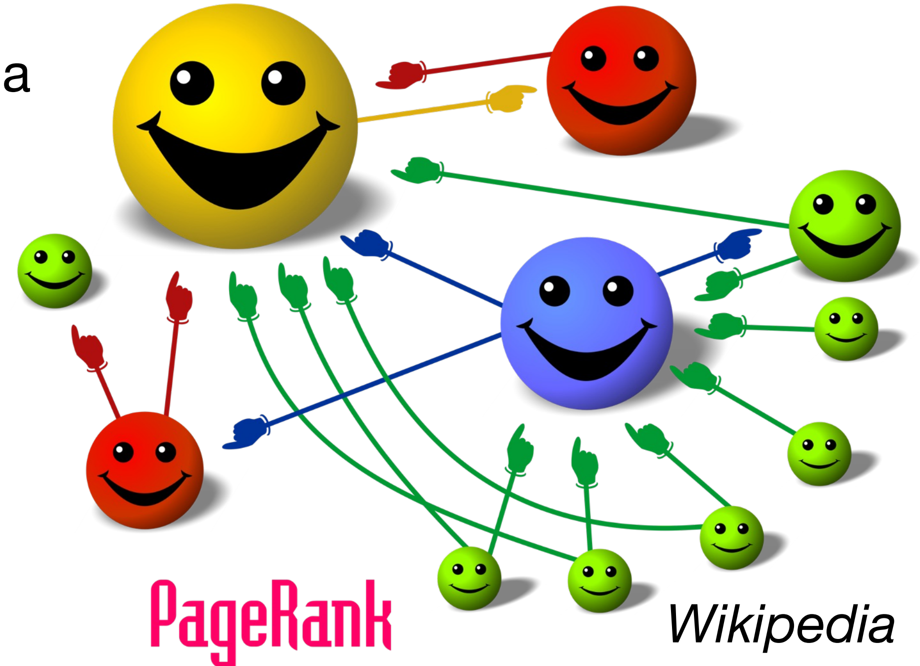
- (Long run) frequency of being in j: π_j
- Frequency of transitions $k \to j$: $\pi_k p_{kj}$
- Frequency of transitions into j: $\sum_k \pi_k p_{kj}$



Example: Google's Pagerank

Consider the following Markov chain:

- One state for each of *m* webpages
- At each time step, choose one of the outgoing links from a page with equal probability
- Rank of a page: Fraction of time that a "random surfer" spends on that page (equilibrium distribution)



Wikipedia

Good Properties:

- Webpages are important if many other pages link to them
- Webpages are important if other important pages link to them

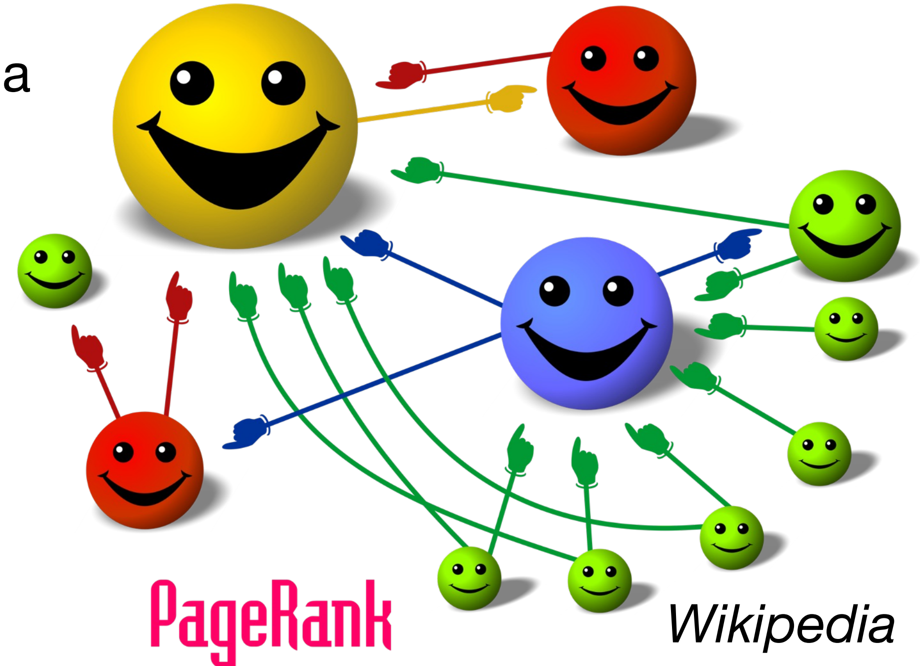
Example: Google's Pagerank

Consider the following Markov chain:

- One state for each of *m* webpages
- At each time step, choose one of the outgoing links from a page with equal probability
- Rank of a page: Fraction of time that a "random surfer" spends on that page (equilibrium distribution)

Problems:

- There may be many absorbing states, so the desired equilibrium distribution does not exist!
- Results can be sensitive to single link changes, and thus "noisy"



Wikipedia

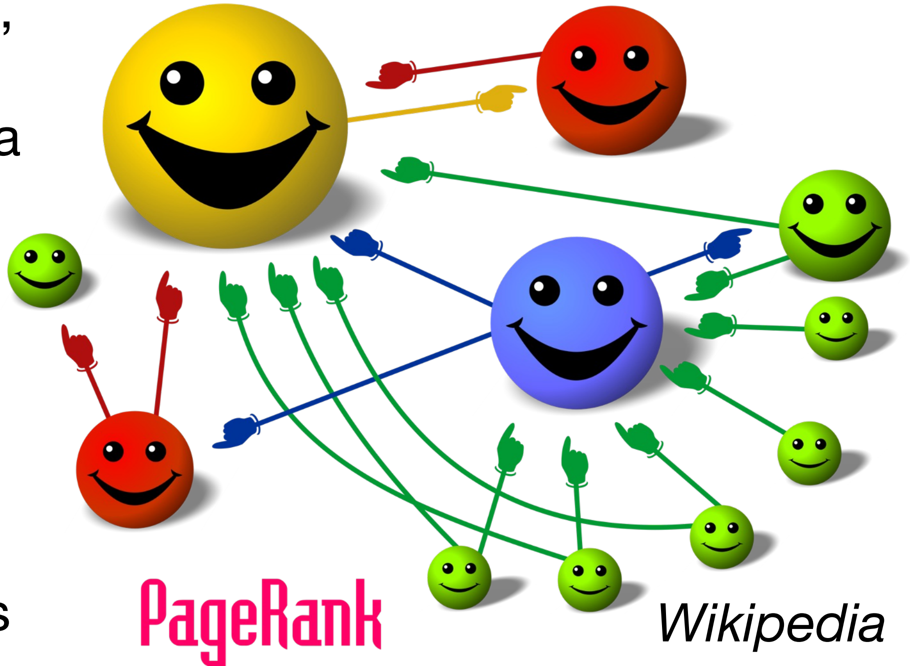
Example: Google's Pagerank

Consider the following Markov chain:

- At each time step, *with probability q*, choose one of the outgoing links from a page with equal probability
- At each time step, *with probability 1-q*, choose one of *m* pages uniformly
- Rank of a page: Fraction of time that a "random surfer" spends on that page

Good Properties:

- Webpages are important if many other pages link to them
- Webpages are important if other important pages link to them
- Steady-state distribution always exists

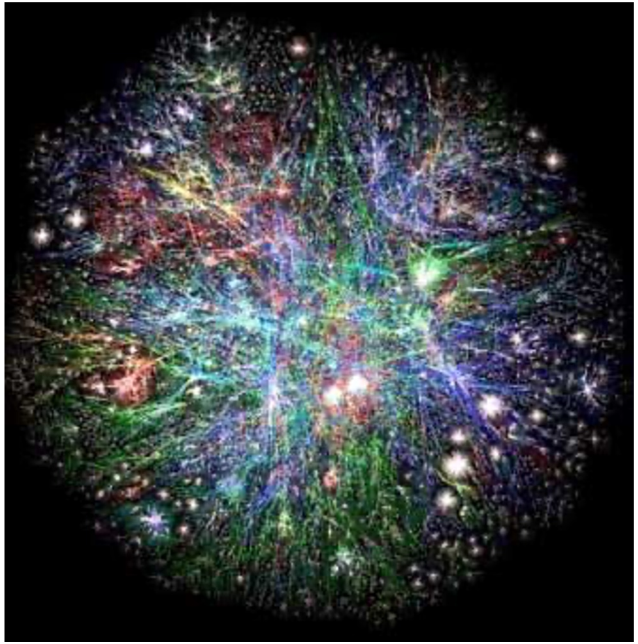


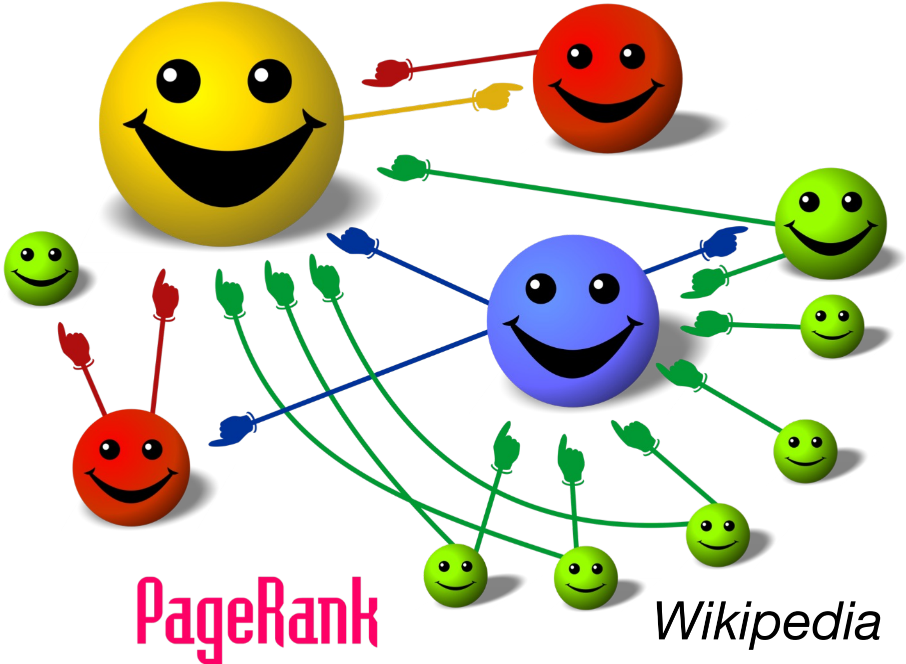
Wikipedia

Example: Google's Pagerank

Because web graph is sparse, it is possible to compute the probabilities of reaching each page even if the number of pages is very large.

$$\pi_n^T = \pi_{n-1}^T P = \pi_{n-2}^T P P = \pi_0^T P^n$$





Wikipedia

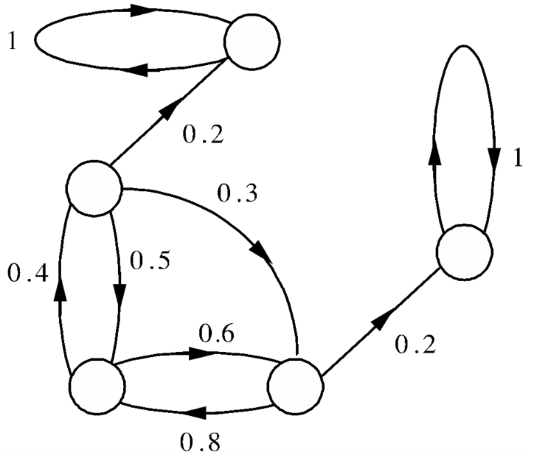
CS145: Lecture 19 Outline

- Markov Chains over Multiple Time Steps
- Markov State Classifications and Durations
- Steady-State Behavior and Equilibrium Distributions
- Absorption in Markov Chains

Transient and Absorbing States

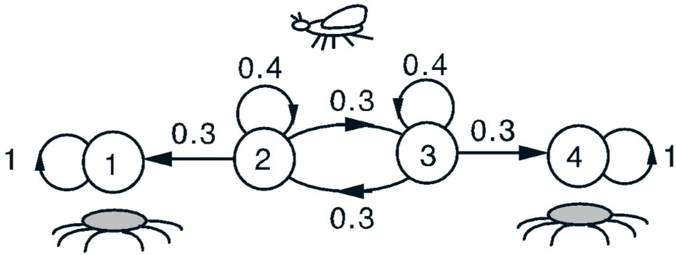
Consider a simplified class of Markov chains where all states are one of two types:

- *Transient:* Visited finite number of times.
- *Absorbing:* Once entered, process never leaves (self-transition probability equals 1).



Practical examples of such Markov chains:

- Games with *win* and *loss* states.
- Computer systems with *failure* states.



To characterize such Markov chains:

- From each *transient* state, what is the probability of each *absorbing* state?
- From each *transient* state, what is the expected time until *absorption*?

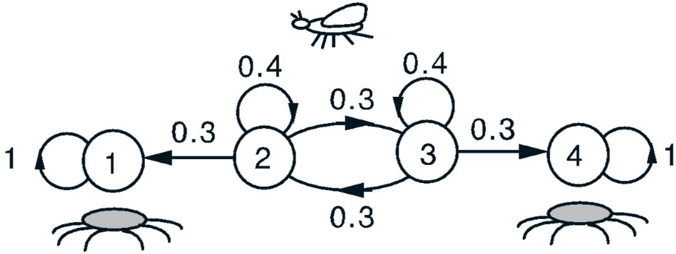
Computing Absorption Probabilities

$a_i = \mathbf{P}(X_n \text{ eventually becomes equal to the absorbing state } s \,|\, X_0 = i)$

Theorem:

Consider a Markov chain in which each state is either transient or absorbing. We fix a particular absorbing state s. Then, the probabilities a_i of eventually reaching state s, starting from i, are the unique solution of the equations

$$a_s = 1,$$
$$a_i = 0, \qquad \text{for all absorbing } i \neq s,$$
$$a_i = \sum_{j=1}^{m} p_{ij} a_j, \qquad \text{for all transient } i.$$



Proof:

$$A = \{\text{state } s \text{ is eventually reached}\}$$

$$\begin{aligned} a_i &= \mathbf{P}(A \,|\, X_0 = i) \\ &= \sum_{j=1}^{m} \mathbf{P}(A \,|\, X_0 = i, X_1 = j)\mathbf{P}(X_1 = j \,|\, X_0 = i) \qquad \text{(total probability thm.)} \\ &= \sum_{j=1}^{m} \mathbf{P}(A \,|\, X_1 = j) p_{ij} \qquad \text{(Markov property)} \\ &= \sum_{j=1}^{m} a_j p_{ij}. \end{aligned}$$

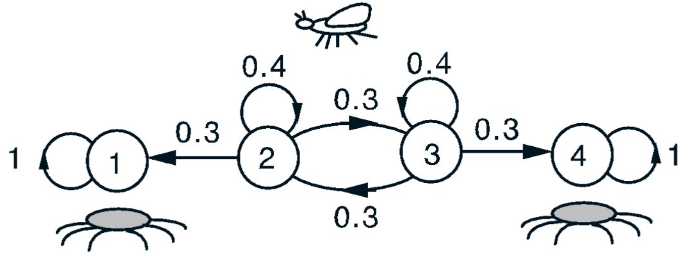
Computing Absorption Probabilities

$a_i = \mathbf{P}(X_n \text{ eventually becomes equal to the absorbing state } s \,|\, X_0 = i)$

Theorem:

Consider a Markov chain in which each state is either transient or absorbing. We fix a particular absorbing state s. Then, the probabilities a_i of eventually reaching state s, starting from i, are the unique solution of the equations

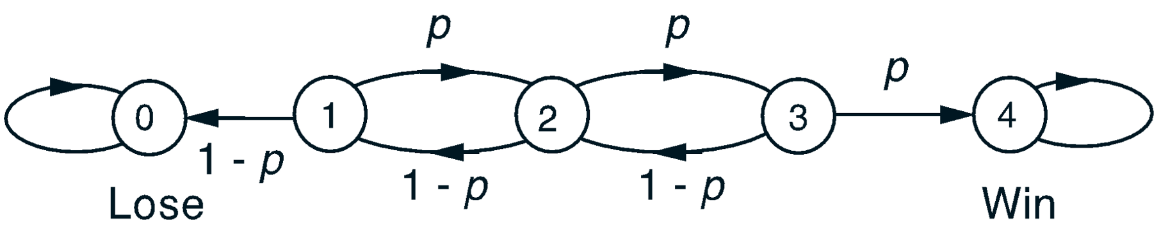
$$
\begin{aligned}
a_s &= 1, \\
a_i &= 0, \qquad \text{for all absorbing } i \neq s, \\
a_i &= \sum_{j=1}^{m} p_{ij} a_j, \qquad \text{for all transient } i.
\end{aligned}
$$



From each initial state i, what is the probability a_i of eventually reaching state 1?

$$a_1 = 1 \qquad a_2 = \frac{2}{3} \qquad a_3 = \frac{1}{3} \qquad a_4 = 0$$

Example: The Gambler's Ruin



$$m = 4$$

- At each round, win \$1 with probability p, lose \$1 with probability $(1-p)$
- Gambler plays until wins a target of \$$m$, or loses all money
- Starting with \$$i$, what is the probability that the gambler *wins* or *loses*?

Let a_i be probability of loss, $1 - a_i$ be probability of win.

$$1 - a_i = \begin{cases} \dfrac{1-\rho^i}{1-\rho^m} & \text{if } \rho \neq 1, \\ \dfrac{i}{m} & \text{if } \rho = 1. \end{cases} \qquad \rho = \frac{1-p}{p}$$

Detailed analysis in B&T Example 7.11.

Expected Time to Absorption

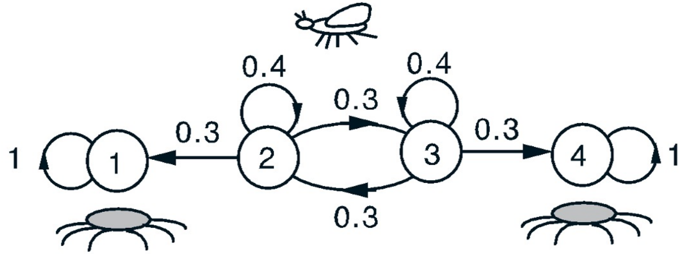
$$\mu_i = \mathbf{E}\big[\text{number of transitions until absorption, starting from } i\big]$$

Theorem:

The expected times μ_i to absorption, starting from state i are the unique solution of the equations

$$\mu_i = 0, \qquad \text{for all recurrent states } i,$$

$$\mu_i = 1 + \sum_{j=1}^{m} p_{ij}\mu_j, \qquad \text{for all transient states } i.$$

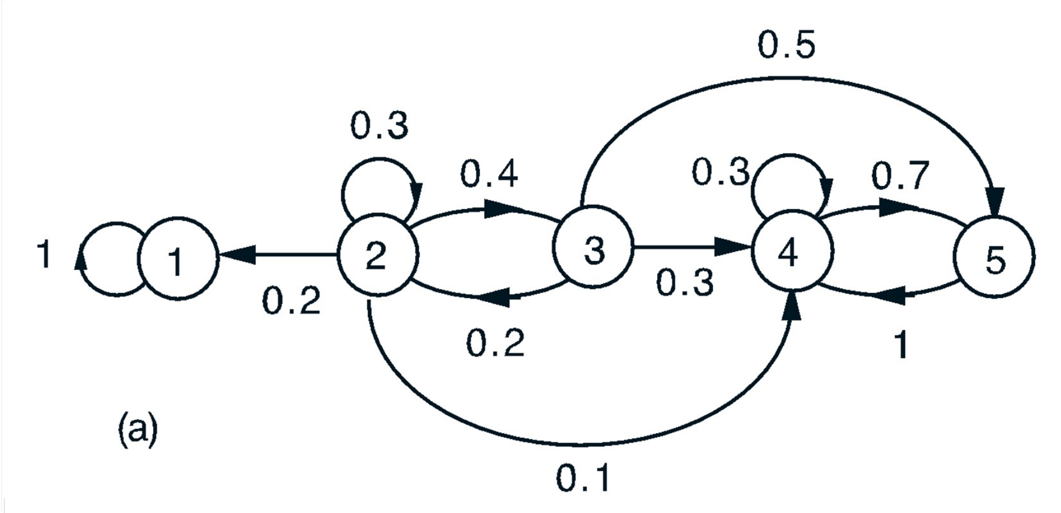


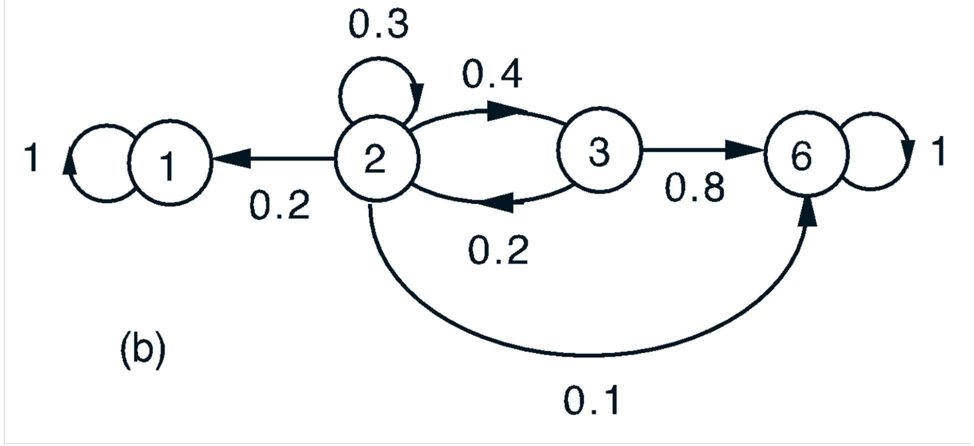
From each initial state *i*, what is the expected number of steps to absorption?

$$\mu_1 = 0 \qquad \mu_2 = \mu_3 = \frac{10}{3} \qquad \mu_4 = 0$$

Matches mean of a geometric distribution with success probability of 0.3.

Absorption in General Markov Chains





- First phase: Find probability of reaching (being absorbed by) each recurrent class
- Second phase: Find equilibrium distribution of states within each of these recurrent classes