AM 121: Introduction to Optimization Models and Methods

Lecture 16: Markov Chains

Yiling Chen
SEAS

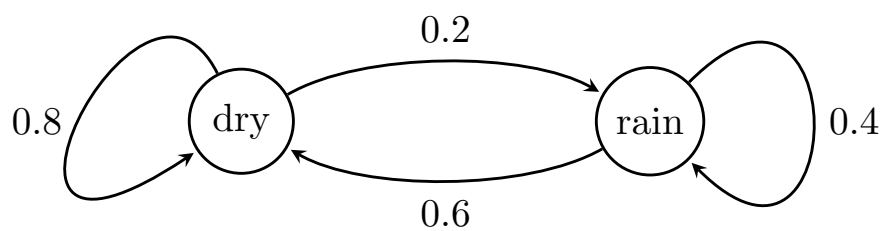
Lesson Plan

- Stochastic process
- Markov Chains
  - n-step probabilities
  - Communicating states, irreducibility
  - Recurrent states, transient states
  - Ergodic states
- Steady-state probabilities
- Applications
  - Google PageRank
  - NCAA prediction

- Jensen & Bard: 11.3, 12, 13

Weather

- Probability 0.8 dry tomorrow if dry today, 0.6 dry tomorrow if rains today. Probabilities do not change if the weather weather before today is taken into account.
- $S_t = \begin{cases} 0 & \text{if day } t \text{ is dry} \\ 1 & \text{if day } t \text{ is rainy} \end{cases}$
- $P(S_{t+1} = 0 | S_t = 0) = 0.8$
- $P(S_{t+1} = 0 | S_t = 1) = 0.6$



State Transition Diagram

Stochastic processes

- A stochastic process is a sequence of random variables $\{S_t\}$, in each period $t \in \{0, 1, 2, \dots\}$
- There is a set of possible states

$$S = \{0, \dots, m-1\}$$

- State $S_t \in S$
- In general, the transition probability is

$$P(S_{t+1} = j | S_0 = i_0, S_1 = i_1, \dots, S_t = i)$$

Markov Chain

- A stochastic process has the Markovian property if

$$P(S_{t+1} = j | S_0 = i_0, S_i = i_1, \ldots, S_t = i) = P(S_{t+1} = j | S_t = i)$$

- Transition probabilities are stationary if

$$P(S_{t+1} = j | S_t = i) = p_{ij}, \qquad \forall\ t, i, j$$

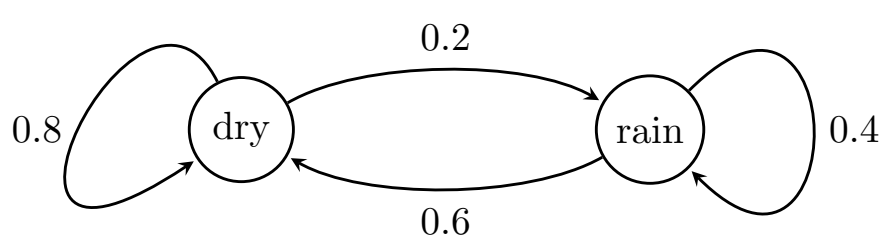
- **Markovian + stationary == Markov chain**

Weather

- Probability 0.8 dry tomorrow if dry today, 0.6 if rains today. Probabilities do not change if weather before today is taken into account.
- $S_t = \begin{cases} 0 & \text{if day } t \text{ is dry} \\ 1 & \text{if day } t \text{ is rainy} \end{cases}$
- $P(S_{t+1} = 0 | S_t = 0) = 0.8$
- $P(S_{t+1} = 0 | S_t = 1) = 0.6$

Transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix}$$



State Transition Diagram

Transition Matrix

$$\mathbf{P} = \begin{array}{c} \\ \\ \\ \text{from } m-1 \rightarrow \end{array} \begin{array}{c} \text{to } 1 \downarrow \\ \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0\ m-1} \\ p_{10} & p_{11} & \cdots & p_{1\ m-1} \\ \cdots & & \cdots & \\ p_{m-1\ 0} & p_{m-1\ 1} & \cdots & p_{m-1\ m-1} \end{pmatrix} \end{array}$$

All rows sum to 1.
Every entry between 0 and 1.

- To define a Markov chain:
  - Describe states $S = \{0, \ldots, m-1\}$
  - Define the transition matrix $\mathbf{P}$

Example: IRS audit

- Often have to **increase the state space** to achieve the Markovian property
- E.g., if the probability of an IRS tax audit depends on the last two audits then we can introduce four states:
- $S = \left\{ \begin{array}{ll} 0 & :: \text{ if last two audits were “no no”} \\ 1 & :: \text{ if last two audits were “yes no”} \\ 2 & :: \text{ if last two audits were “no yes”} \\ 3 & :: \text{ if last two audits were “yes yes”} \end{array} \right\}$

n-step Transition Probability

- n-step transition probability $p_{ij}^{(n)}$ is the **conditional probability** of state j after n steps from state i

$$P(S_{t+n} = j | S_t = i) = p_{ij}^{(n)}$$

- $p_{ij}^{(n)} = \sum_{k=0}^{m-1} p_{ik}^{(d)} p_{kj}^{(n-d)}$
  - i.e., in going from i to j in n steps, must be in some state k after d steps and j in $n-d$ steps; sum over all possible intermediate states
- For example (given two states):

$$p_{01}^{(2)} = p_{00}p_{01} + p_{01}p_{11}$$

$$p_{11}^{(2)} = p_{10}p_{01} + p_{11}p_{11}$$

n-step Transition Probability

- n-step transition probability $p_{ij}^{(n)}$ is the **conditional probability** of state j after n steps from state i

$$P(S_{t+n} = j | S_t = i) = p_{ij}^{(n)}$$

- We have:

$$\mathbf{P}^{(2)} = \mathbf{P}\mathbf{P} = \mathbf{P}^2$$

$$\mathbf{P}^{(3)} = \mathbf{P}\mathbf{P}^{(2)} = \mathbf{P}\mathbf{P}^2 = \mathbf{P}^3$$

$$\mathbf{P}^{(4)} = \mathbf{P}\mathbf{P}^{(3)} = \mathbf{P}\mathbf{P}^3 = \mathbf{P}^4$$

Example: Weather

$$\mathbf{P}^{(2)} = \mathbf{PP} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{pmatrix}$$

$$\mathbf{P}^{(3)} = \mathbf{PP}^{(2)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{pmatrix} = \begin{pmatrix} 0.752 & 0.248 \\ 0.744 & 0.256 \end{pmatrix}$$

$$\mathbf{P}^{(4)} = \mathbf{PP}^{(3)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 0.752 & 0.248 \\ 0.744 & 0.256 \end{pmatrix} = \begin{pmatrix} 0.75 & 0.25 \\ 0.749 & 0.251 \end{pmatrix}$$

$$\mathbf{P}^{(5)} = \mathbf{PP}^{(4)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \begin{pmatrix} 0.75 & 0.25 \\ 0.749 & 0.251 \end{pmatrix} = \begin{pmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \end{pmatrix}$$

- After 5 steps, each row has identical entries: the prob of it being dry or rainy is independent of the weather 5 days ago.
- Will see later that 0.75, 0.25 are steady-state probabilities.

Unconditional n-step State Probability

- The unconditional n-step probability $P(S_n = j)$ is the probability that will be in state j after n steps.
- For this, need probability distribution on initial state, i.e. $P(S_0 = i)$ for $i \in S$. Given this,

$$P(S_n = j) = P(S_0 = 0)p_{0j}^{(n)} + \cdots + P(S_0 = m-1)p_{m-1\ j}^{(n)}$$

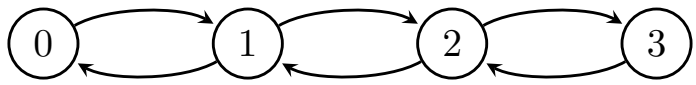
- **Weather example**: if dry on day one ($S_0 = 0$) with probability 0.5, then probability dry after two days is

$$P(S_2 = 0) = 0.5p_{00}^{(2)} + 0.5p_{10}^{(2)} = 0.5 \times 0.76 + 0.5 \times 0.72 = 0.74$$

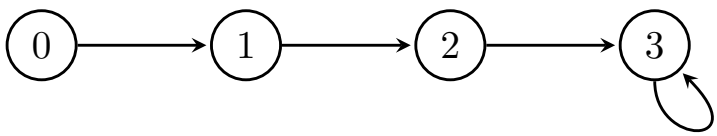
Classes of States

absence of edge = prob zero. Existence of edge = prob > 0.

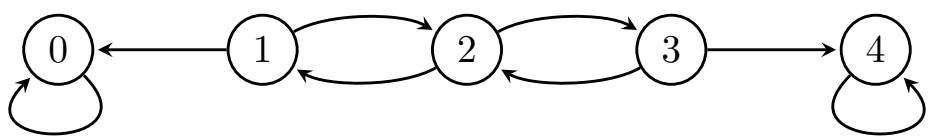
"communicate" = can transition from i to j and from j to n (after some number of steps)
One class in which all states communicate (irreducible):



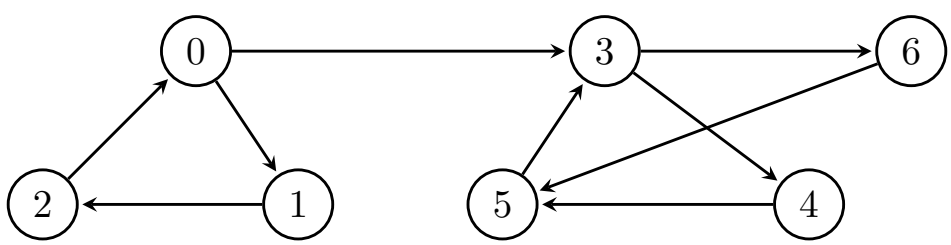
Four distinct classes, no states communicate (except with themselves):



More Examples



Class 1 Class 2 Class 3



Class 1 Class 2

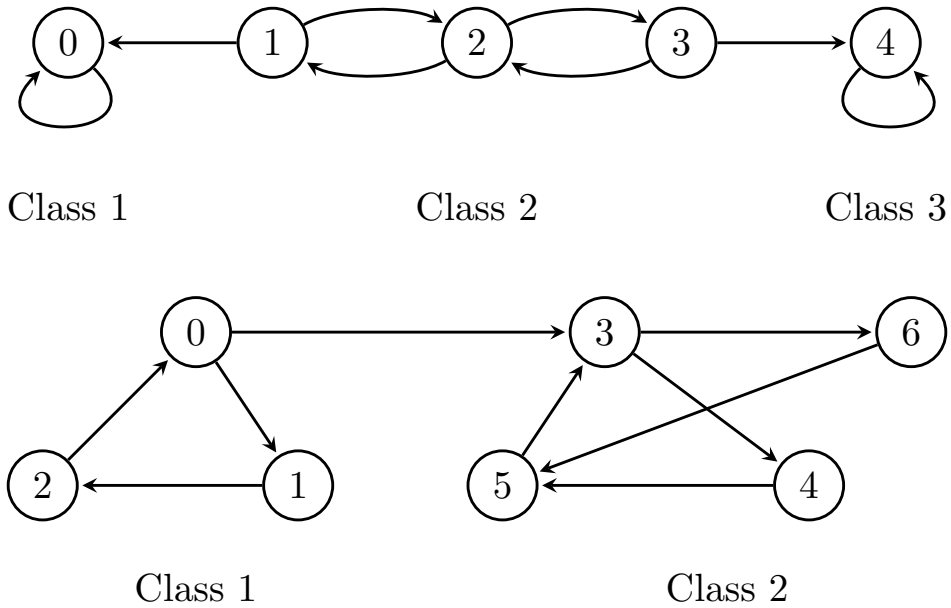
Classes of States; Irreducible

- State j is accessible from state i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$
- States i and j communicate if state j is accessible from i and state i from state j
  - any state communicates with itself
  - transitive (if i communicates with j, and j with k, then i communicates with k)
- A class is a **set of states that communicate with each other and is maximal** (cannot add another state to the set)
- A Markov chain is irreducible if it has only one class.

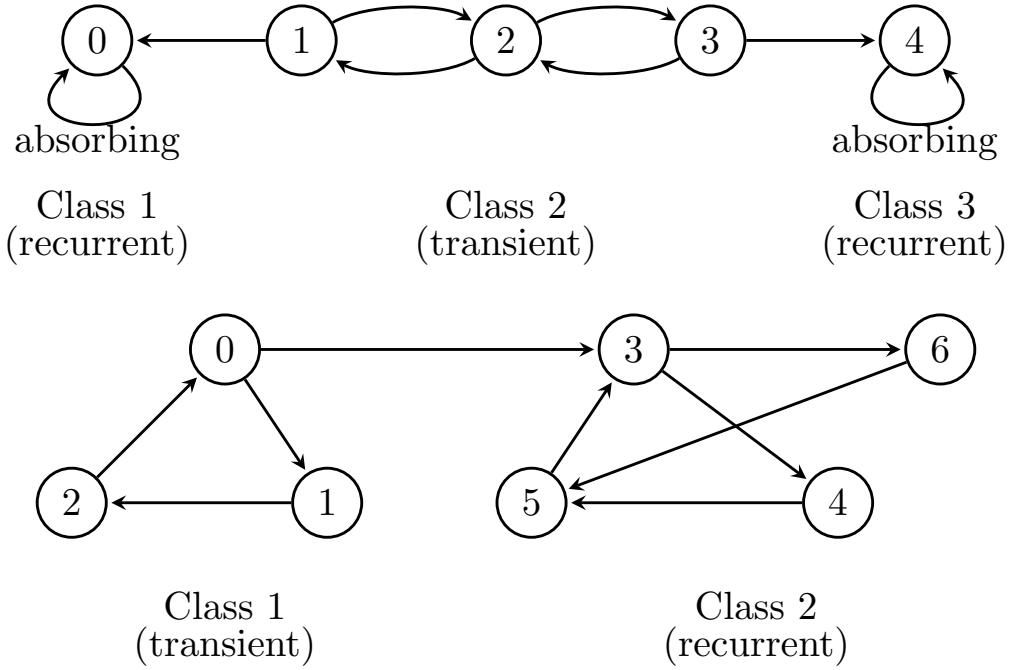
Recurrent and Transient States

- A state is recurrent if the process will return to the state again with probability $= 1$. A state is transient otherwise.
- **Proposition.** Recurrence is a class property: all states in a class are either recurrent or transient.
- **Proposition.** Not all states can be transient.
  $\Rightarrow$ all states in an irreducible Markov chain are recurrent.
- A state is absorbing if the process will never leave the state.

Examples

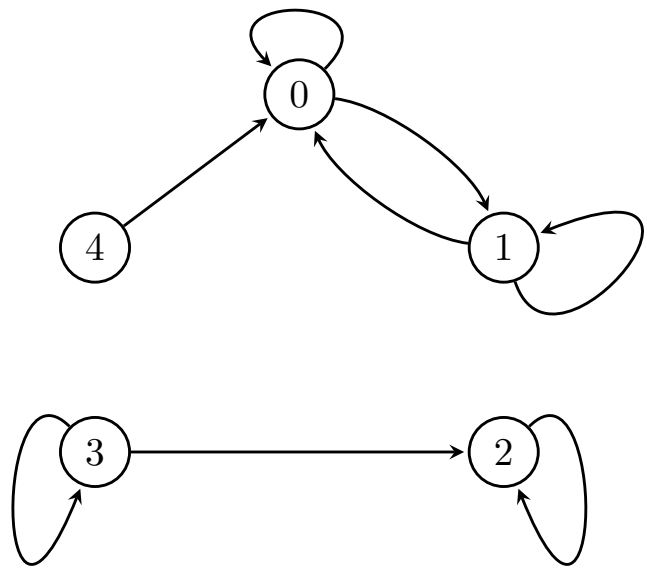


Examples



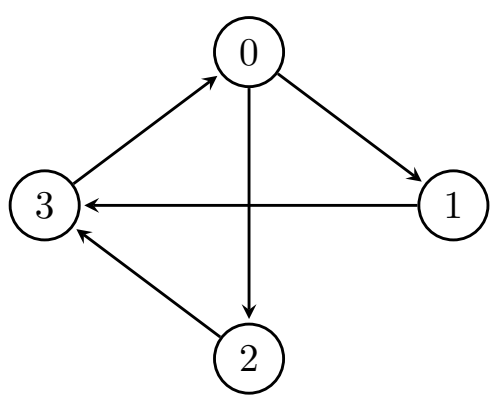
Note: recurrent classes have no leaving arcs

Another Example (Exercise)



Periodicity

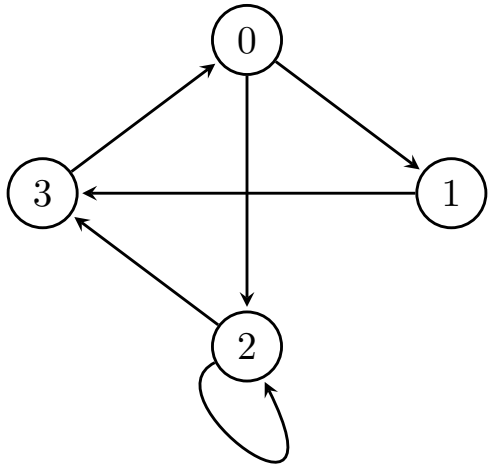
- The period of state i, written $d(i)$, is an integer $d(i) \geq 1$ such that the chain **can return to i only at multiples of the period $d(i)$ and $d(i)$ is the largest such integer.**
- **Proposition.** Periodicity is a class property: all states in a class have the same period.
- Example:



irreducible (necessarily recurrent)
every state is periodic with period 3

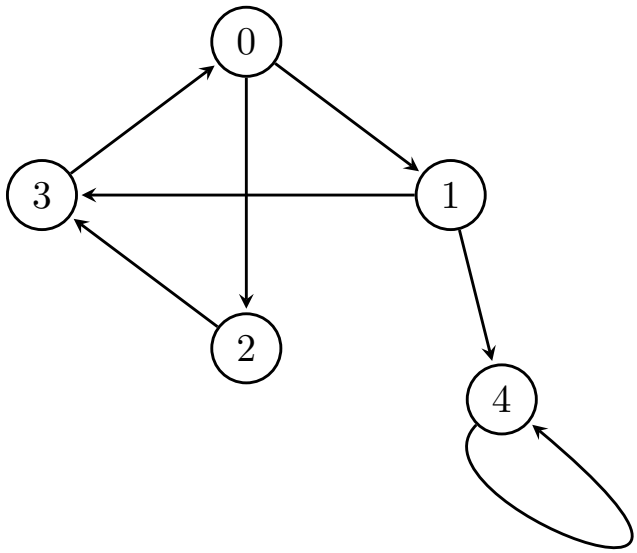
Aperiodic states

- State i is aperiodic if the period is $d(i) = 1$ and periodic otherwise.
  - It is sufficient for $p_{ii} > 0$ for state i to be aperiodic.
  - A state to which the chain can never return is defined to be aperiodic.
  - *Remember: periodicity or aperiodicity is a class property*

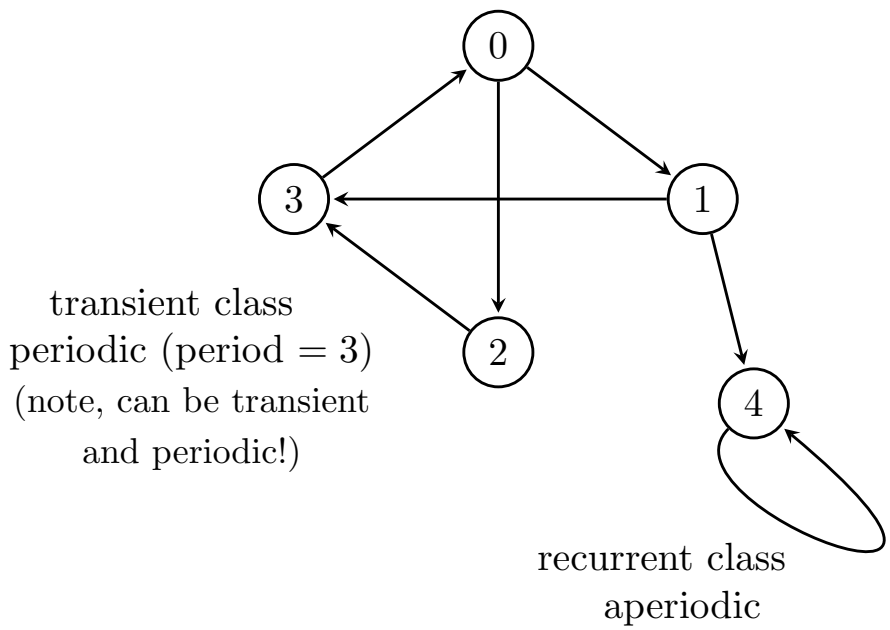


irreducible (necessarily recurrent)
all states are aperiodic (**why?**)

Another Example

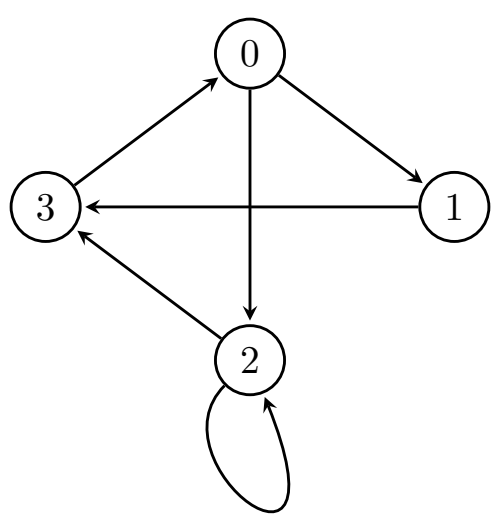


Another Example



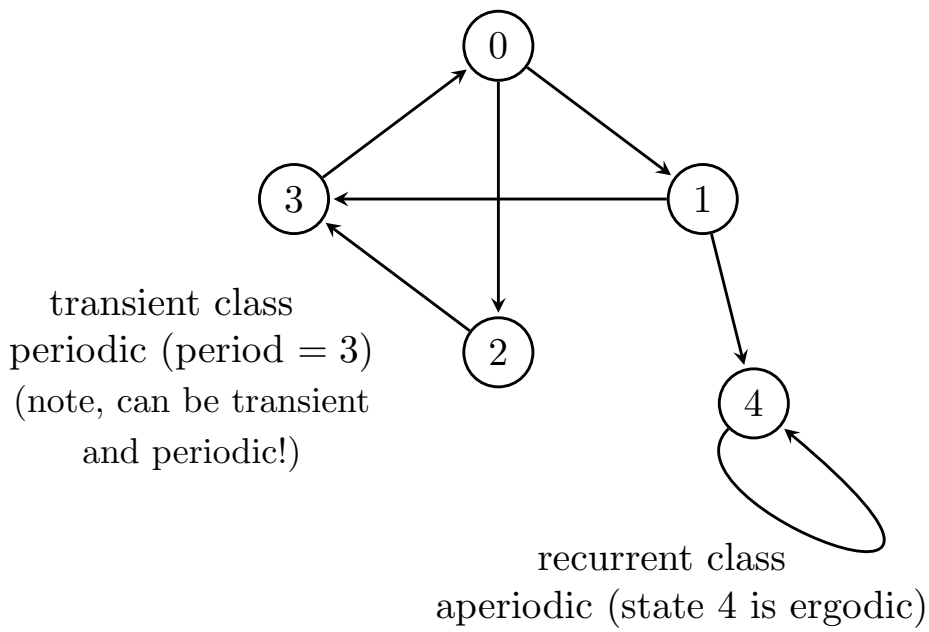
Ergodicity

- An ergodic state is a state that is recurrent and aperiodic.
- An irreducible (thus, recurrent) Markov chain is ergodic if it is aperiodic.



all states are aperiodic and recurrent; an ergodic chain

Example (A Non Ergodic Chain)



Steady-State Probabilities

- For any ergodic Markov chain, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j > 0$ exists, and is independent of i.
- The steady-state probability π_j for state j is the probability the process is in state j after a large number of transitions.
- Computed via the steady-state equations ("balance equations"):

$$\pi_j = \sum_{i=0}^{m-1} \pi_i \cdot p_{ij} \quad \text{for } j = 0, \ldots, m-1 \tag{1}$$

$$\sum_{j=0}^{m-1} \pi_j = 1 \tag{2}$$

Example: Weather

- Ergodic: Can solve for steady-state probabilities
- $\pi_0 = \pi_0 p_{00} + \pi_1 p_{10} = 0.8\pi_0 + 0.6\pi_1$
- $\pi_1 = \pi_0 p_{01} + \pi_1 p_{11} = 0.2\pi_0 + 0.4\pi_1$
- $1 = \pi_0 + \pi_1$
- Get: $\pi_0 = 0.75$, $\pi_1 = 0.25$
- Steady state probability dry 0.75, steady state probability rain 0.25

Computing the Steady State Probabilities

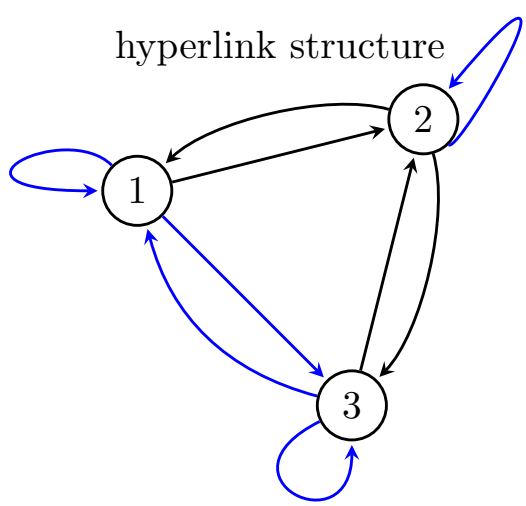
- $\pi_j = \sum_{i=0}^{m-1} \pi_i \cdot p_{ij}$ for $j = 0, \ldots, m-1$
- $\sum_{j=0}^{m-1} \pi_j = 1$
- $m+1$ equations, m unknowns. One of first m equations is redundant.
- We need the $\sum_j \pi_j = 1$ equation because otherwise $\pi_j = 0$ solves.
- Can solve via Gaussian elimination, matrix inversion, etc.

Application: Google PageRank

- Link from page 1 to page 2 indicates that 1 "trusts" 2.
- The quality of a page is related to its in-degree as well as the quality of the incoming pages.
- Random surfer model: with prob $1 - \alpha$ follow an out-edge at random (e.g., 1/2), w. prob α do a random jump.
- Defines a Markov chain. The PageRank of page i is the steady state probability of visiting page i:
  - the fraction of time the surfer will spend at the page

(Brin and Page'98)

Example: PageRank



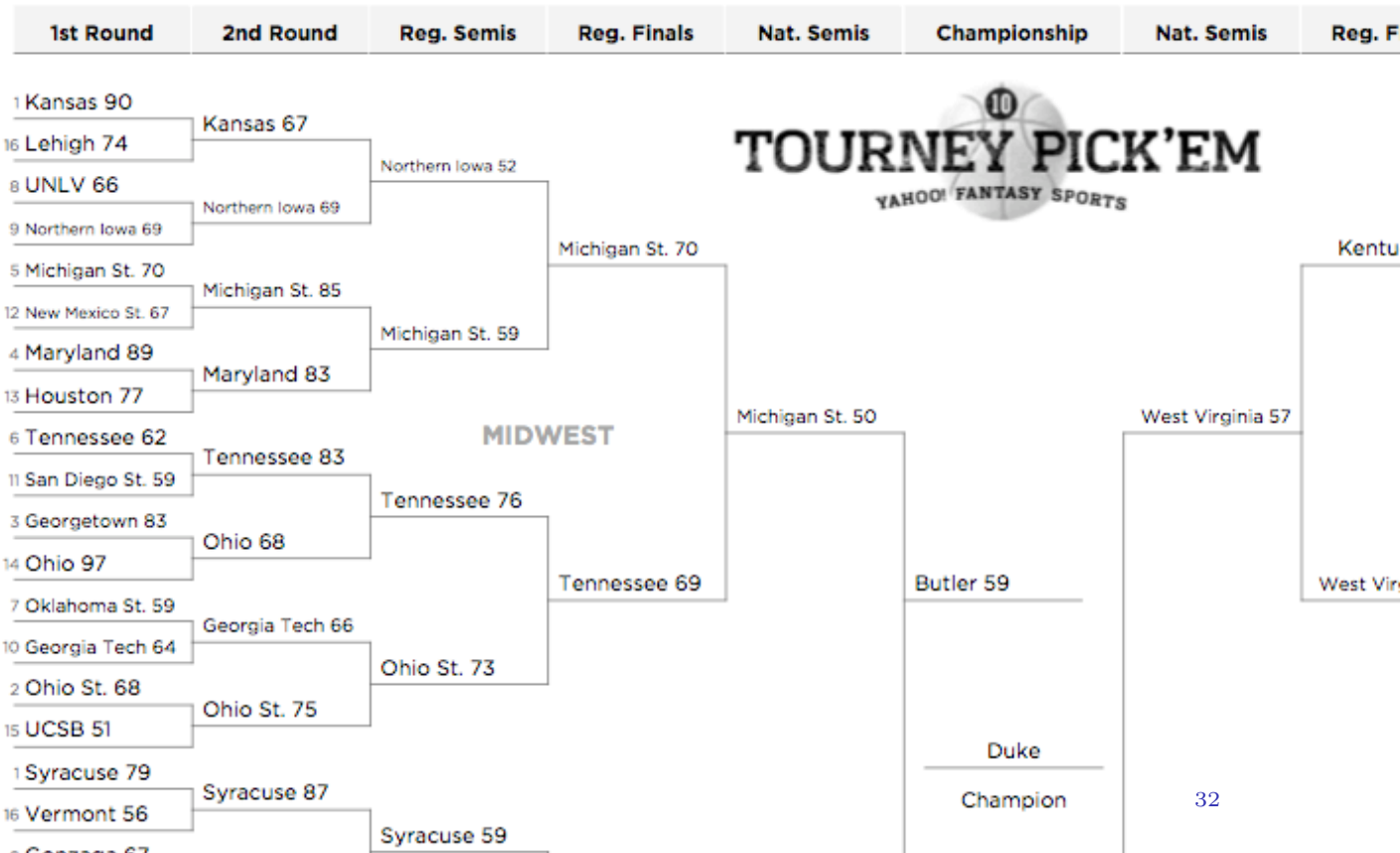
include jump prob 1/6 for missing edges in this example

- $\mathbf{P} = \begin{pmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{pmatrix}$
- $\pi = (5/18, 4/9, 5/18)$, the PageRank (node 2 is the most trusted!)

Application: NCAA Basketball

- FBI estimates that more than $3 billion wagered on NCAA basketball each year. Betting pools typically set-up to predict the winner of each game.
- Season is mid-Nov to mid-March, followed by a 64-team tournament (31 champions from the conferences, plus the remaining best teams).
- Seeded into 4 regions. Within each region play single-elimination tournament on neutral court.
- The winner of each region goes to the Final Four; play a single-elimination tournament on neutral court.

http://sports.yahoo.com/ncaab/bracket



Many Sources of Information

- AP poll of sportswriters
- ESPN/USA Today poll of coaches
- Sagarin ratings (USA Today)
- Massey ratings
- Las Vegas odds

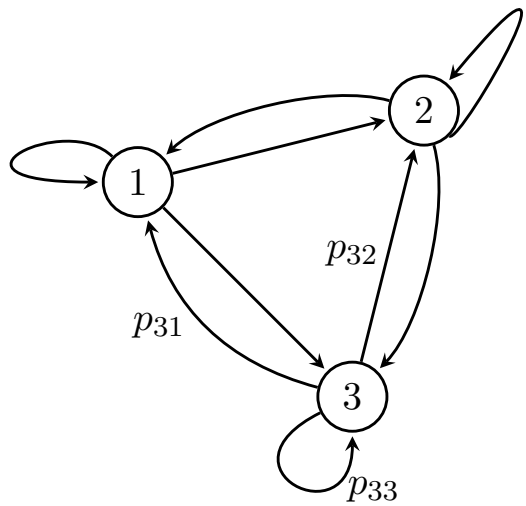
- Can Markov chains be used to do better than all of these?
  - P. Kvam and J. S. Sokol, "A logistic regression/Markov chain model for NCAA basketball," *Naval Research Logistics* **53**: 788-803 (2006).

The NCAA Data

- 330 Division I teams
- Members i and j in a conference play each other twice a year, once on i's home court and once on j's home court
- Data is the location, winner, and scores.

- Idea: Build a Markov chain to represent the random walk of an expert who transitions according to the team she thinks is best ("follows the wins").

Example: Three NCAA Teams



- Random surfer model: Current state i is the team that the expert believes is best. At each step, pick a competitor j at random, and toss a weighted coin to determime whether j wins (transition to j) or not (stay at state i.)
- Steady-state probabilities give estimated quality of each team

Methodology: NCAA via Markov Chains

1. Each game is ordered pair $(i, j) \in G$, (set of games G) visiting team first.
2. z_{ij} is the win margin of home team j (> 0 if wins)
3. Use a logistic regression to estimate $r(z_{ij})$: the prob home team j would beat i at neutral site
4. Uses $r(z_{ij})$ values to compute transition probabilities p_{ij}
5. From this, compute steady-state probabilities of random walk; higher probability of visiting a team = better team.

Defining the Transition Probabilities

- p_{ij}: transition probability from i to j:

$$p_{ij} \propto \left(\sum_{j:(i,j)\in G} r(z_{ij}) + \sum_{j:(j,i)\in G} (1 - r(z_{ji})) \right)$$

i, j games where j at home

prob j beats i given margin z_{ij} at home

i, j games where i at home

prob i beats j given margin z_{ji} at home

- Define $p_{ii} = 1 - \sum_{j \neq i} p_{ij}$

- Use four years of NCAA data (1999-2000 to 2002-2003 seasons)

S_x: Given that team 1 margin over team 2 is x points on 1's home court, what is the prob that 1 beats 2 on 2's home court?

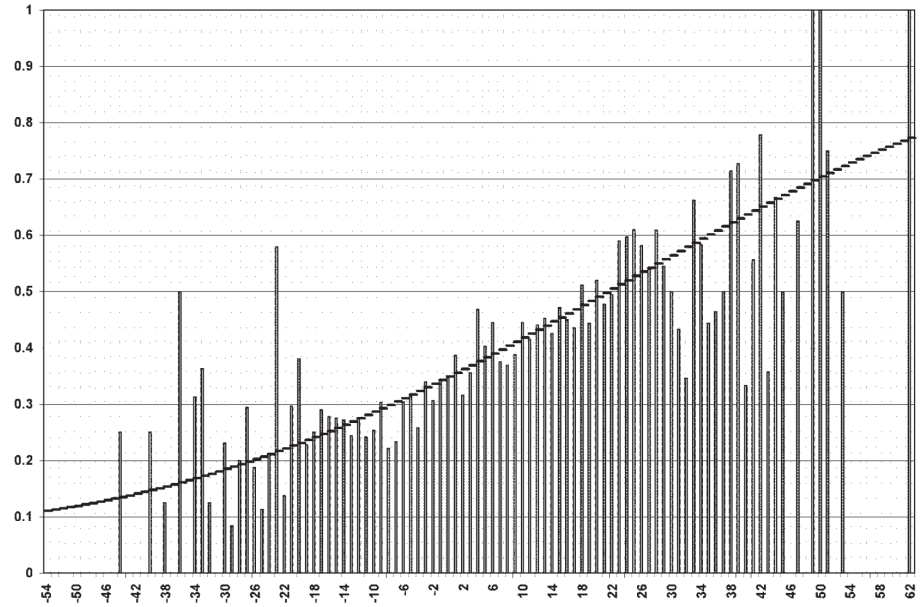


Figure 3. Observed values and logistic regression estimates for S_x^H.

Estimate home-court advantage h as $h := x/2$, with $x \approx 22$, so that $S_x = 0.5$. Given this, estimate $r(z) := S_{z+h}$.

Obs. prob. home team j wins (i, j) based on $\pi_j - \pi_i$

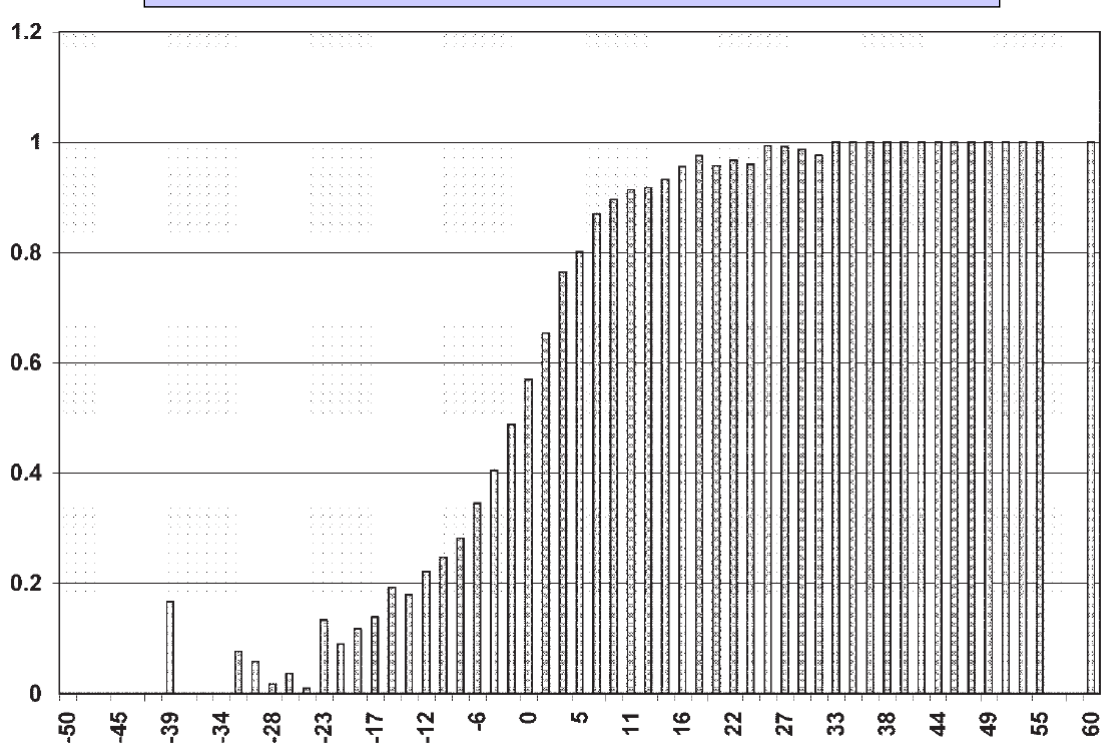
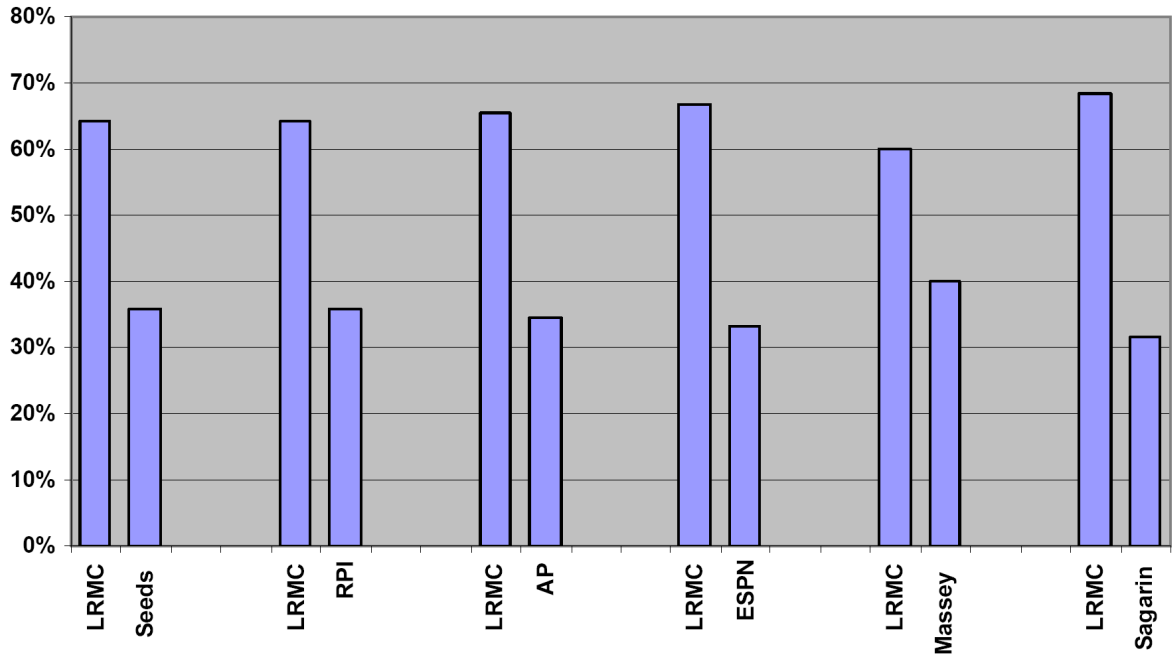
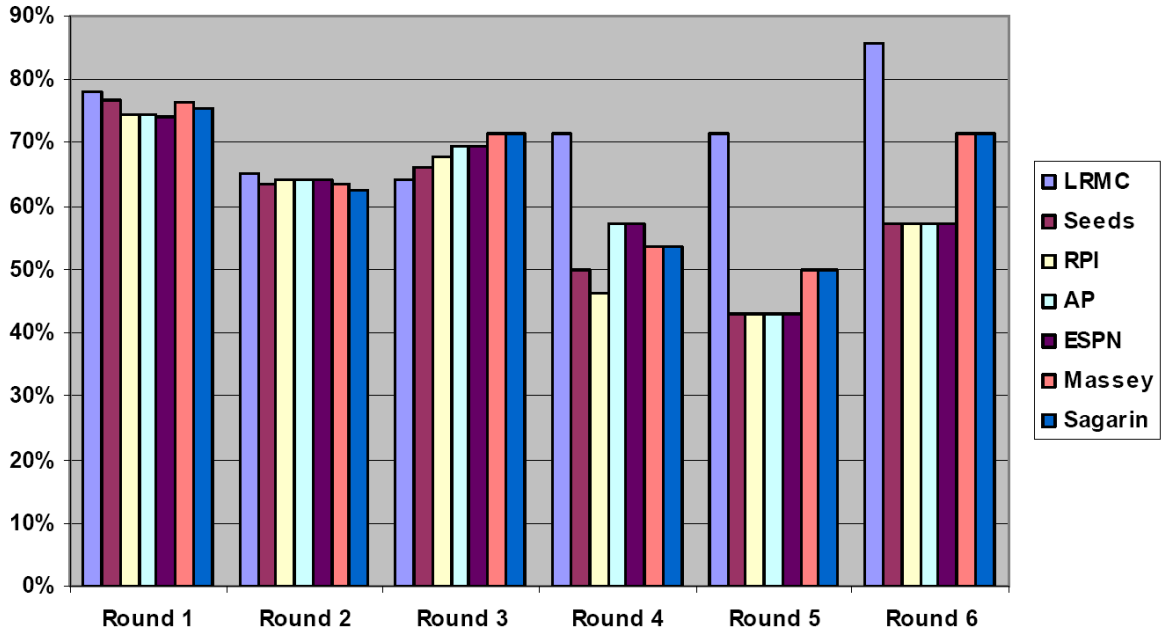


Figure 5. Observed win probability by steady state difference $\times\ 10^{-4}$.



Head-to-head prediction results (LRMC vs. others) in games where predictions disagree.



Percentage of correct predictions in each round, 2000-06. LRMC is highest in 5 of the 6 rounds, with large advantages in rounds 4-6.

Summary

- A Markov chain is Markovian and stationary.
- A class of states communicate with each other; a chain is irreducible if one class
- Class properties:
  - Recurrent (return w.p. 1) vs transient
  - Periodic (only return after multiples of $d > 1$) vs aperiodic
- A state is ergodic if it is recurrent and aperiodic. An irreducible Markov chain is ergodic if it is aperiodic.
- Steady-state probabilities are well-defined in an ergodic Markov chain.