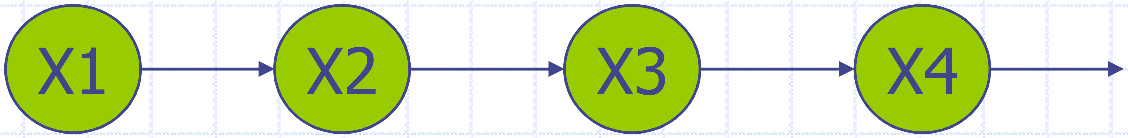
# Markov Models

# Markov Chain

- A sequence of states: $X_1, X_2, X_3, \ldots$
  - Usually over time
- The transition from $X_{t-1}$ to $X_t$ depends only on $X_{t-1}$ (Markov Property).
  - A Bayesian network that forms a chain
  - The transition probabilities are the same for any t (stationary process)



Courtsey of Michael Littman

# Example: Gambler's Ruin

- Specification:
  - Gambler has 3 dollars.
  - Win a dollar with prob. 1/3.
  - Lose a dollar with prob. 2/3.
  - Fail: no dollars.
  - Succeed: Have 5 dollars.
- States: the amount of money
  - 0, 1, 2, 3, 4, 5
- Transition Probabilities

# Transition Probabilities

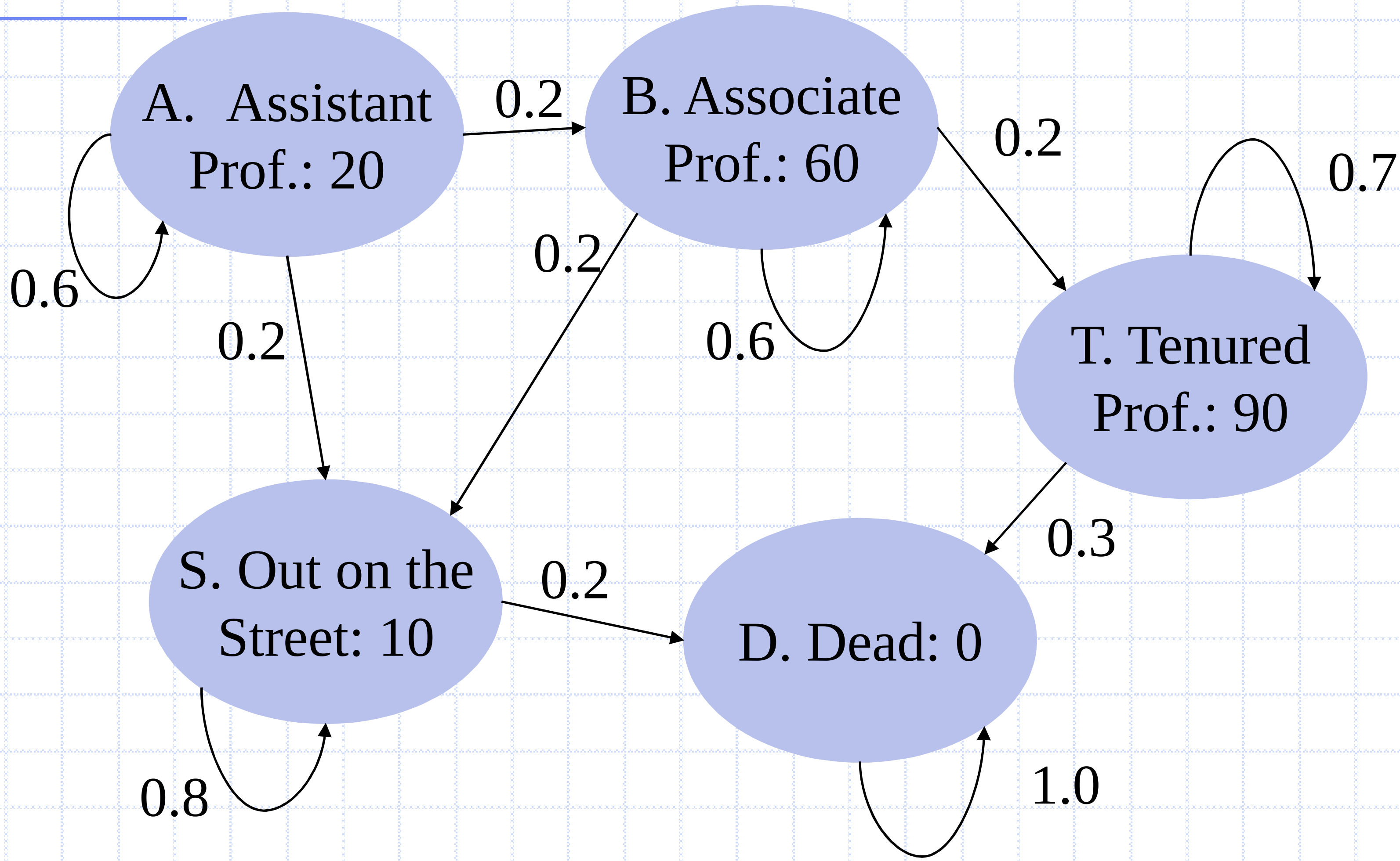
- Suppose a state has N possible values
  - $X_t=s_1$, $X_t=s_2$,….., $X_t=s_N$.
- $N^2$ Transition Probabilities
  - $P(X_t=s_i|X_{t-1}=s_j)$, $1\le i, j \le N$
- The transition probabilities can be represented as a NxN matrix or a directed graph.
- Example: Gambler's Ruin

# What can Markov Chains Do?

- Example: Gambler's Ruin
  - The probability of a particular sequence
    - 3, 4, 3, 2, 3, 2, 1, 0
  - The probability of success for the gambler
  - The average number of bets the gambler will make.

Courtsey of Michael Littman

# Example: Academic Life




What is the expected lifetime income of an academic?

# Solving for Total Reward

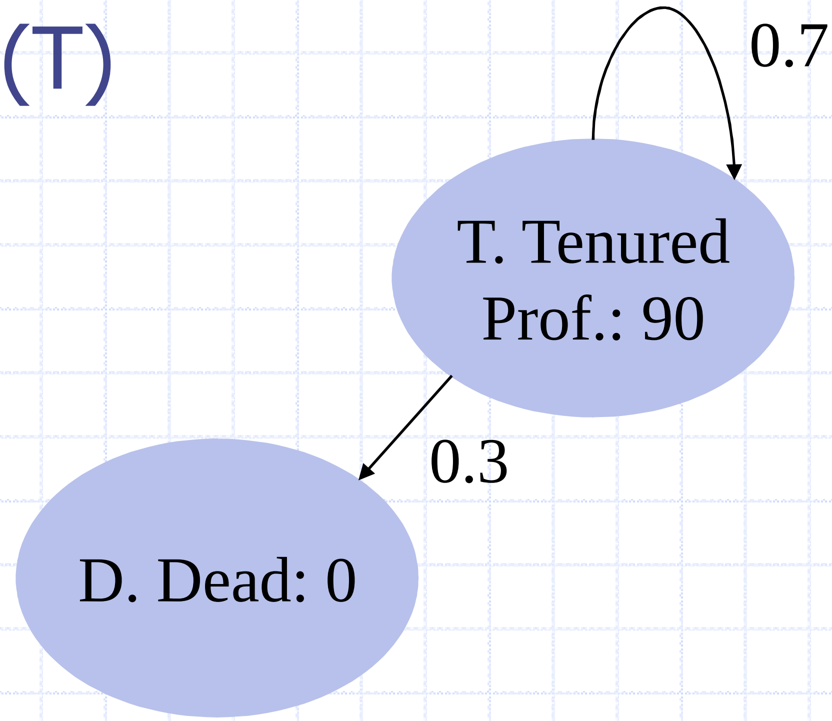
- L(i) is expected total reward received starting in state i.
- How could we compute L(A)?
- Would it help to compute L(B), L(T), L(S), and L(D) also?

# Solving the Academic Life

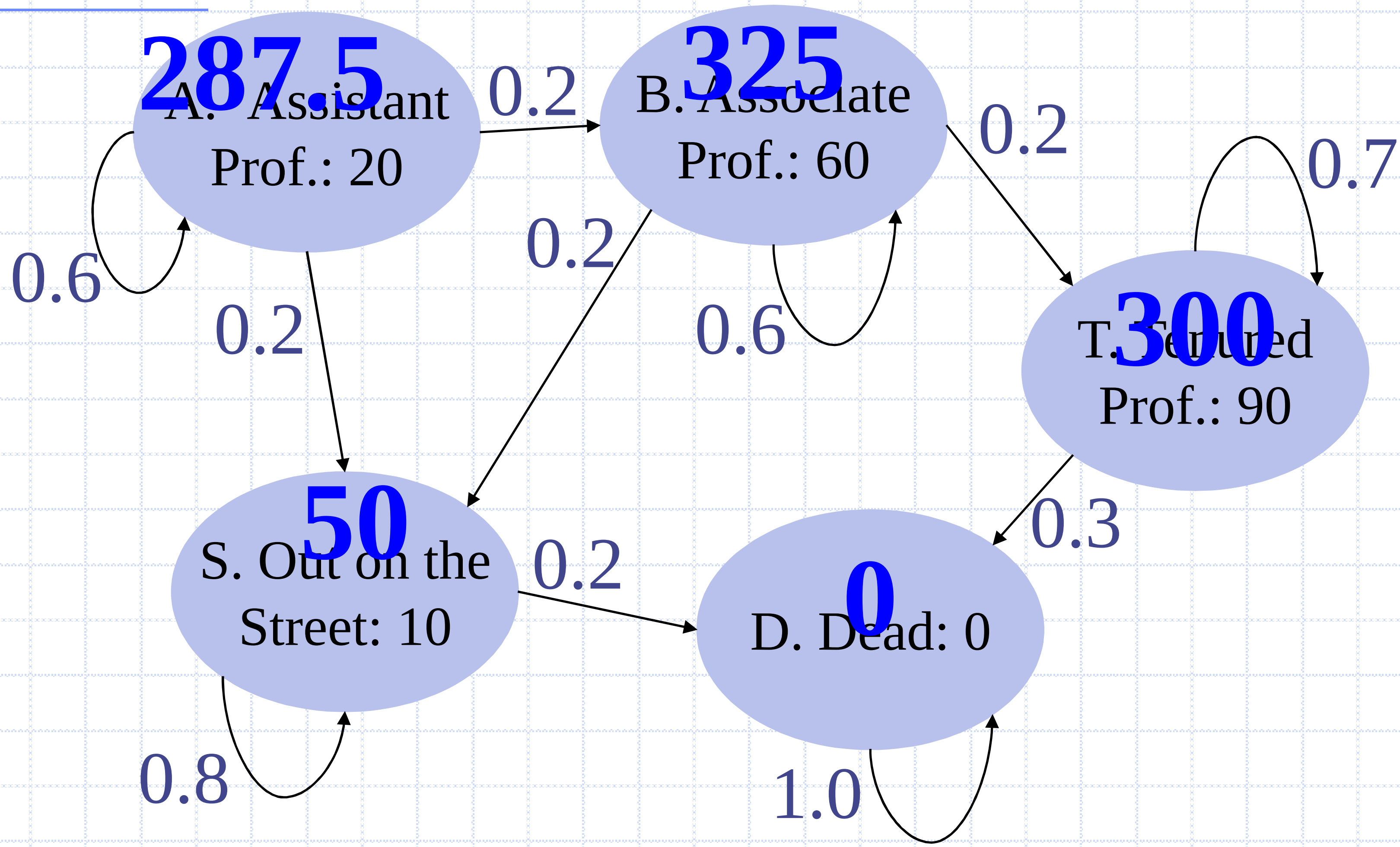
- The expected income at state D is 0
- $L(T)=90+0.7x90+0.7^2x90+\ldots$

  $L(T)=90+0.7xL(T)$

  $L(T)=300$

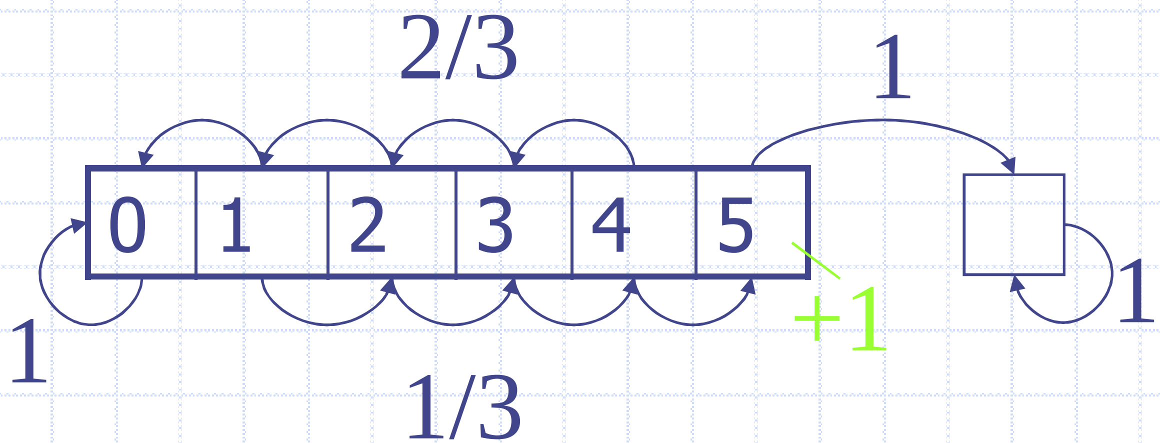


# Working Backwards

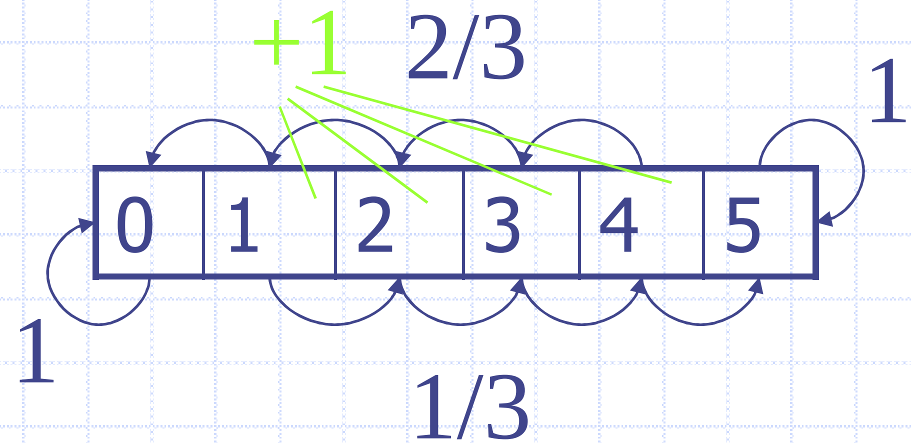



Another question: What is the life expectancy of professors?

# Ruin Chain



# Gambling Time Chain



# Google's Search Engine

- Assumption: A **link** from page A to page B is a **recommendation** of page B by the author of A (we say B is *successor* of A)

➔ Quality of a page is related to its in-degree

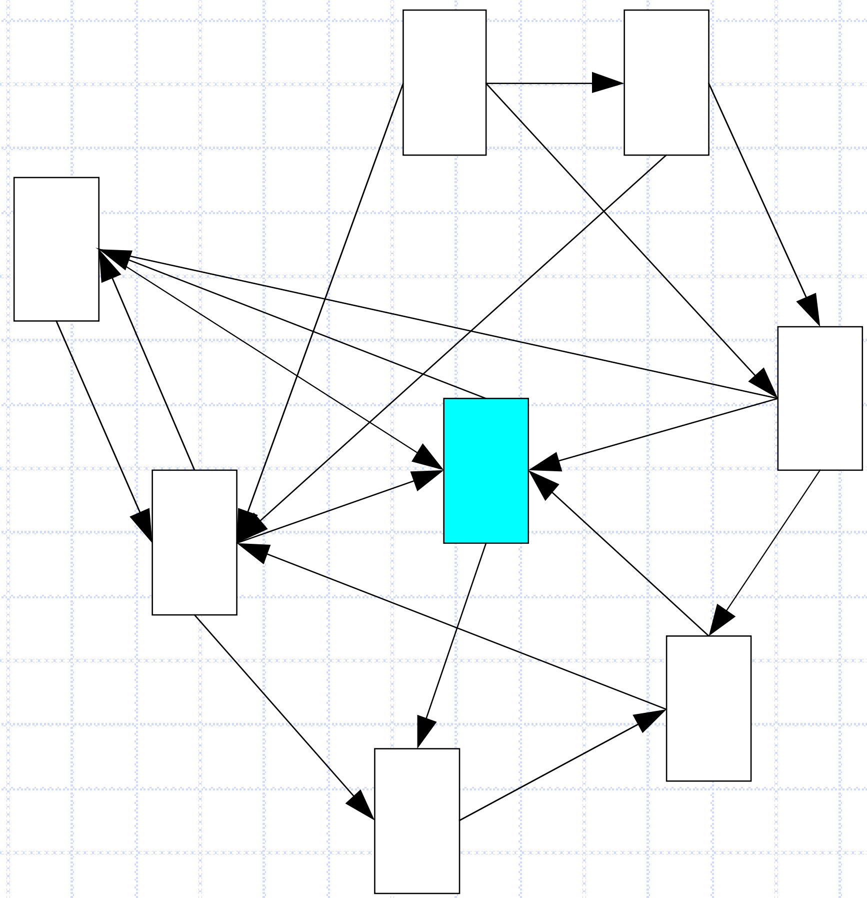
- Recursion: Quality of a page is related to
  - its in-degree, and to
  - the *quality* of pages linking to it

➔ **PageRank** [Brin and Page '98]

# Definition of PageRank

- Consider the following infinite random walk (surf):
  - Initially the surfer is at a random page
  - At each step, the surfer proceeds
    - to a randomly chosen web page with probability d
    - to a randomly chosen successor of the current page with probability 1-d
- **The PageRank of a page p is *the fraction of steps the surfer spends at p in the limit.***

# Random Web Surfer



What’s the probability of a page being visited?

# Stationary Distributions

- Let
  - S is the set of states in a Markov Chain
  - P is its transition probability matrix
- The initial state chosen according to some probability distribution $q^{(0)}$ over S
- $q^{(t)}$ = row vector whose i-th component is the probability that the chain is in state i at time t
- $q^{(t+1)} = q^{(t)} P \;\Rightarrow q^{(t)} = q^{(0)} P^t$
- A stationary distribution is a probability distribution q such that $q = q\,P$ (steady-state behavior)

# Markov Chains

- Theorem: Under certain conditions:
  - There exists a unique stationary distribution q with $q_i > 0$ for all i
  - Let N(i,t) be the number of times the Markov chain visits state i in t steps. Then,

$$\lim_{t \to \infty} \frac{N(i,t)}{t} = q_i$$

# PageRank

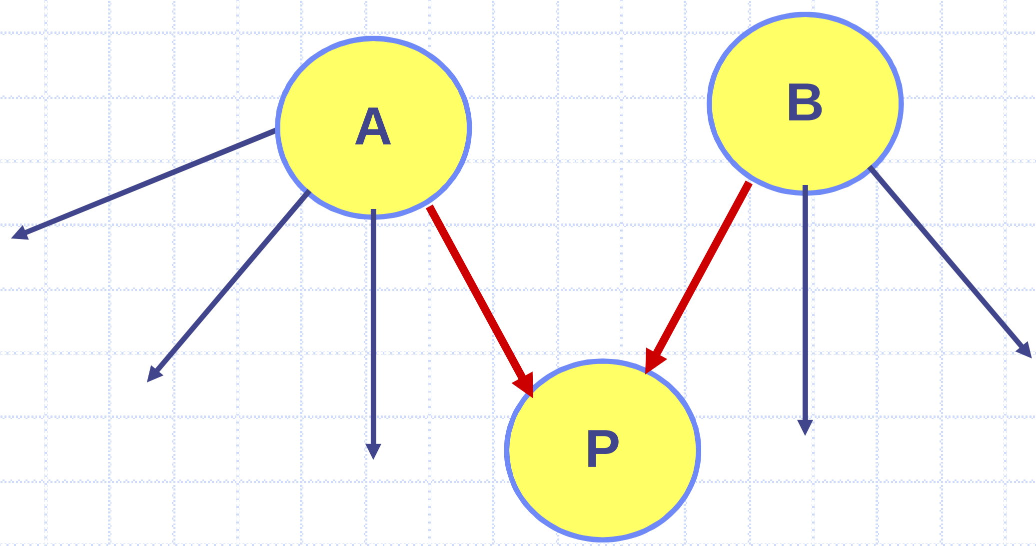
- PageRank = the probability for this Markov chain, i.e.

$$PageRank(u) = \frac{d}{n} + (1-d) \sum_{(v,u)\in E} PageRank(v) / outdegree(v)$$

  where n is the total number of nodes in the graph
  d is the probability of making a random jump.

- Query-independent
- Summarizes the “web opinion” of the page importance

# PageRank

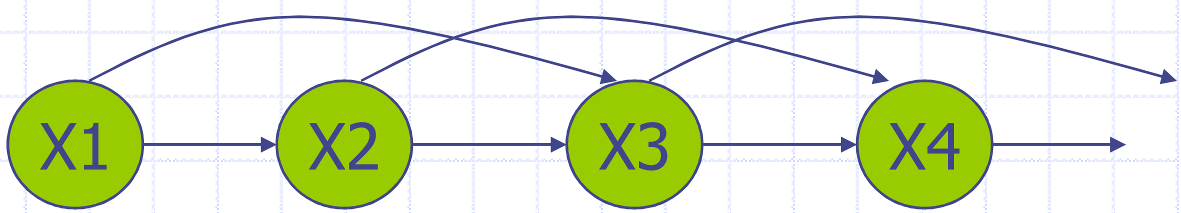



PageRank of P is

$(1\text{-}d)*\left(\ 1/4^{th}\ \text{the PageRank of A} + 1/3^{rd}\ \text{the PageRank of B}\right) + d/n$

# Kth-Order Markov Chain

- What we have discussed so far is the first-order Markov Chain.
- More generally, in kth-order Markov Chain, each state transition depends on previous k states.
  - What's the size of transition probability matrix?

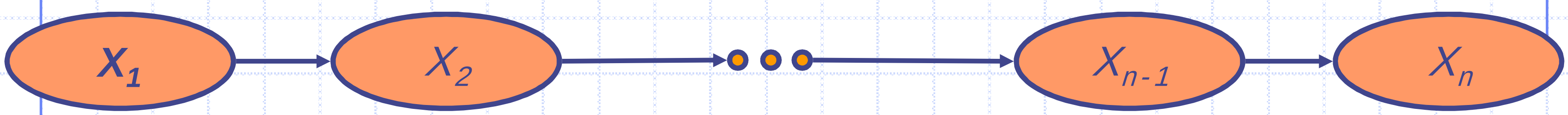


# Finite Markov Chain

An ***integer time stochastic process***, consisting of a ***domain D*** of $m>1$ states $\{s_1,\ldots,s_m\}$ and

1. An ***m*** dimensional ***initial distribution vector*** $(p(s_1),\ldots, p(s_m))$.
2. An ***m×m transition probabilities matrix M***= $(a_{s_i s_j})$

# Markov Chain (cont.)

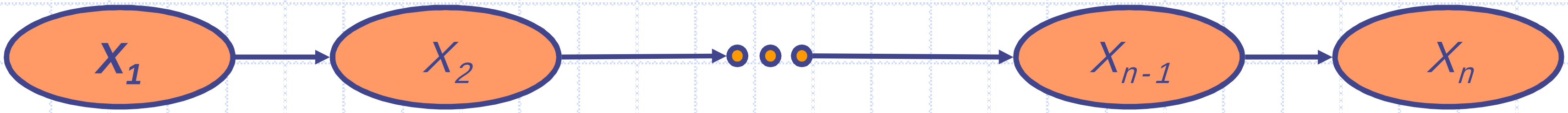



- For each integer *n*, a Markov Chain assigns probability to sequences $(x_1 \ldots x_n)$ over ***D*** (i.e, $x_i \in$ ***D***) as follows:

$$p((x_1, x_2, \ldots x_n)) = p(X_1 = x_1) \prod_{i=2}^{n} p(X_i = x_i \mid X_{i-1} = x_{i-1})$$

$$= p(x_1) \prod_{i=2}^{n} a_{x_{i-1} x_i}$$

# Markov Chain (cont.)




Similarly, each $\boldsymbol{X_i}$ is a probability distributions over $\boldsymbol{D}$, which is determined by the initial distribution $(p_1,..,p_n)$ and the transition matrix $\boldsymbol{M}$.

There is a rich theory which studies the properties of such "Markov sequences" $(\boldsymbol{X_1,\dots, X_i},\dots)$. A bit of this theory is presented next.

## Slide 22

1 this slide was separated from the previous one _after_ the lecture at fall05-6,

, 12/3/2005

# Matrix Representation

| | A | B | C | D |
|---|---|---|---|---|
| A | 0.95 | 0 | 0.05 | 0 |
| B | 0.2 | 0.5 | 0 | 0.3 |
| C | 0 | 0.2 | 0 | 0.8 |
| D | 0 | 0 | 1 | 0 |

The transition probabilities Matrix ***M*** $=(a_{st})$

***M*** is a stochastic Matrix:

$$\sum_t a_{st} = 1$$

The initial ***distribution vector*** $(u_1 \ldots u_m)$ defines the distribution of $X_1$ $(p(X_1 = s_i) = u_i)$ .

Then after one move, the distribution is changed to $X_2 = X_1 M$

# Matrix Representation

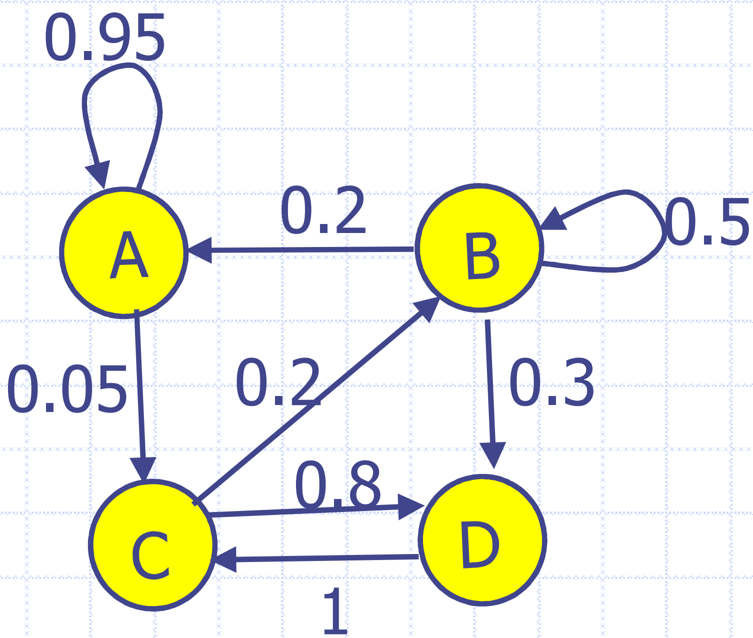
| | A | B | C | D |
|---|---|---|---|---|
| A | 0.95 | 0 | 0.05 | 0 |
| B | 0.2 | 0.5 | 0 | 0.3 |
| C | 0 | 0.2 | 0 | 0.8 |
| D | 0 | 0 | 1 | 0 |

Example: if $\mathbf{X_1}$=(0, 1, 0, 0) then $\mathbf{X_2}$=(0.2, 0.5, 0, 0.3)

And if $\mathbf{X_1}$=(0, 0, 0.5, 0.5) then $\mathbf{X_2}$=(0, 0.1, 0.5, 0.4).

The $i$-th distribution is $X_i = X_1 M^{i-1}$

# Representation of a Markov Chain as a Digraph



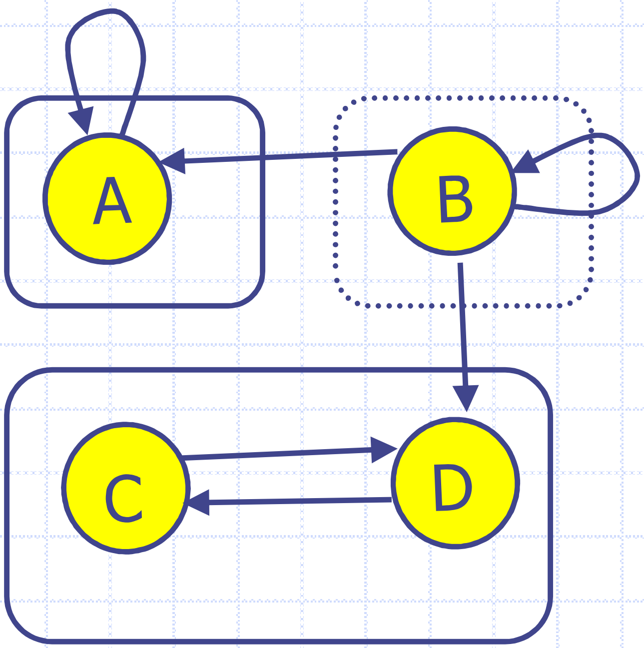

| | A | B | C | D |
|---|---|---|---|---|
| A | 0.95 | 0 | 0.05 | 0 |
| B | 0.2 | 0.5 | 0 | 0.3 |
| C | 0 | 0.2 | 0 | 0.8 |
| D | 0 | 0 | 1 | 0 |

Each directed edge A→B is associated with the **positive** transition probability from A to B.

# Properties of Markov Chain states

- States of Markov chains are classified by the digraph representation (omitting the actual probability values)

- A, C and D are ***recurrent*** states: they are in strongly connected components which are **sinks** in the graph.

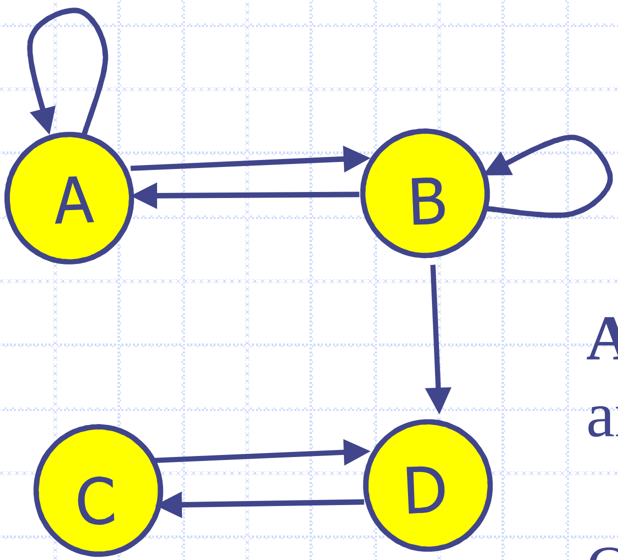



- B is not recurrent – it is a ***transient*** state

Alternative definitions:

- A state ***s*** is ***recurrent*** if it can be reached from any state reachable from ***s;*** otherwise it is ***transient***.

# Another example of Recurrent and Transient States

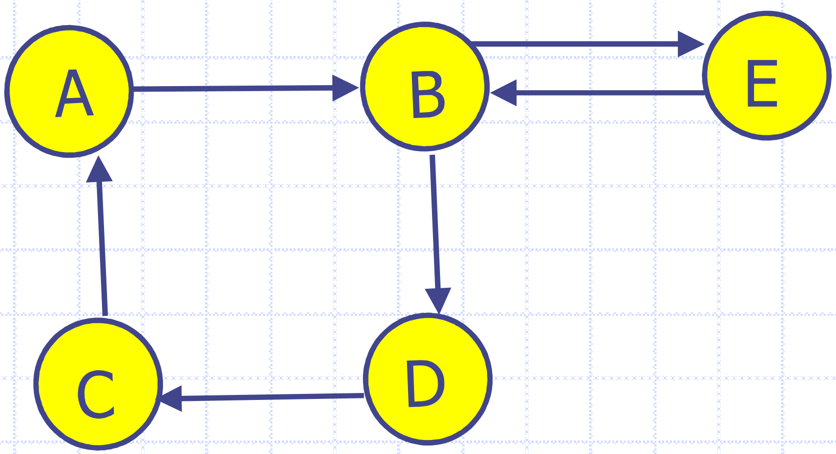


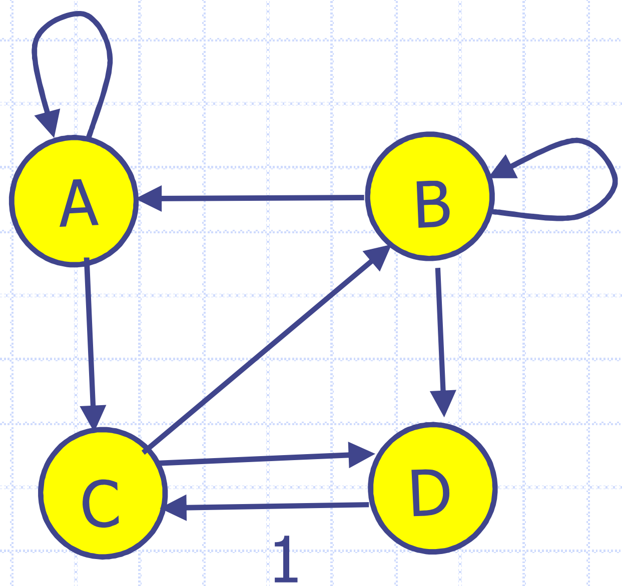

**A** and **B** are ***transient*** states**,** **C** and **D** are ***recurrent*** states.

Once the process moves from **B** to **D**, it will never come back.

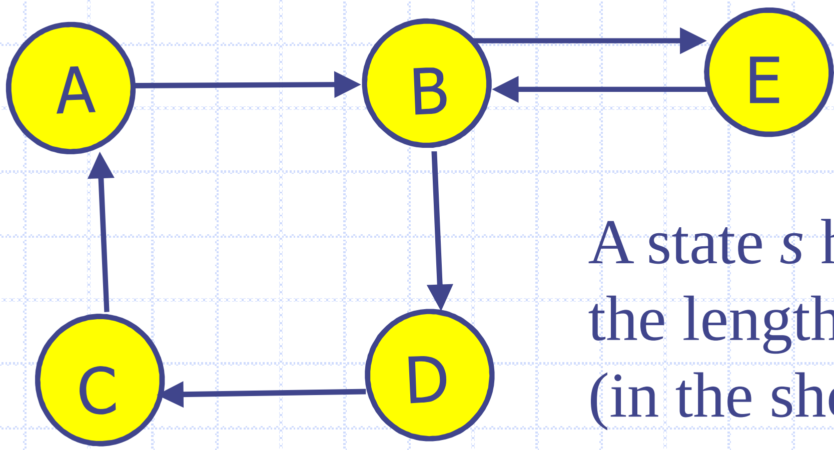
# Irreducible Markov Chains

- A Markov Chain is ***irreducible*** if the corresponding graph is strongly connected (and thus all its states are recurrent).



# Periodic States

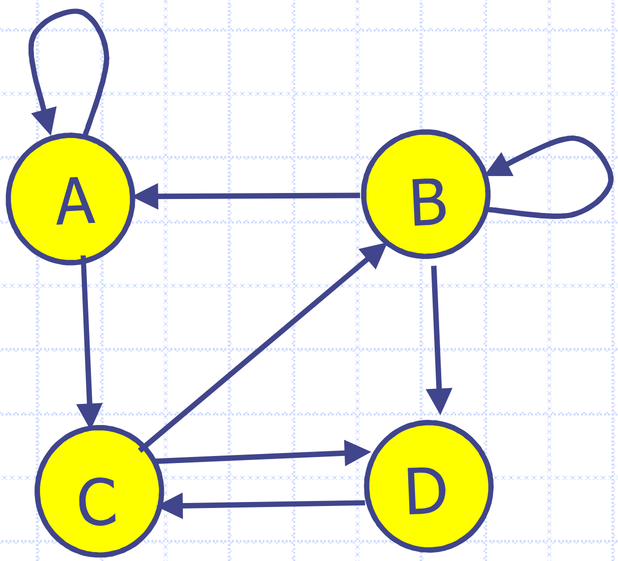



A state $s$ has a period $\boldsymbol{k}$ if $\boldsymbol{k}$ is the ***GCD*** of the lengths of all the cycles that pass via $s$**.** (in the shown graph the period of A is 2).

Exercise: All the states in the same strongly connected component have the same period

A Markov Chain is ***periodic*** if all the states in it have a period $k$ >1. It is ***aperiodic*** otherwise.

# Ergodic Markov Chains




A Markov chain is ***ergodic*** if :

1. ***the corresponding graph is strongly connected.***
2. ***It is not peridoic***

Ergodic Markov Chains are important since they guarantee the corresponding Markovian process converges to a unique distribution, in which all states have strictly positive probability.

# Stationary Distributions for Markov Chains

Let $M$ be a Markov Chain of $m$ states, and let $V = (v_1,\ldots,v_m)$ be a probability distribution over the $m$ states

***V*** $=$ $(v_1,\ldots,v_m)$ is ***stationary distribution*** for ***M*** if ***VM=V.***
(ie, if one step of the process does not change the distribution).

$\Updownarrow$

***V*** is a stationary distribution

***V*** is a left (row) Eigenvector of ***M*** with Eigenvalue 1.

2 example of stationary vector (on the board):
(0.8, 0.2) where M is:

$$\begin{matrix} 0.75 & 0.25 \\ 1 & 0 \end{matrix}$$

, 11/12/2004

# Stationary Distributions for a Markov Chain

Exercise: A stochastic matrix always has a real left Eigenvector with Eigenvalue 1 (hint: show that a stochastic matrix has a right Eigenvector with Eigenvalue 1. Note that the left Eigenvalues of a Matrix are the same as the right Eiganvlues).

[It can be shown that the above Eigenvector $V$ can be non-negative. Hence each Markov Chain has a stationary distribution.]

# "Good" Markov chains

- A Markov Chains is *good* if the distributions $X_i$, as $i \to \infty$:

(1) converge to a unique distribution, independent of the initial distribution.

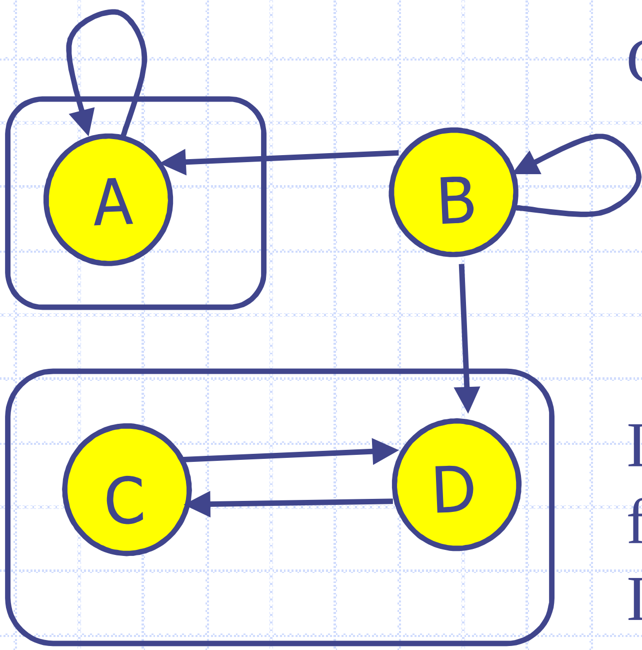
(2) In that unique distribution, each state has a positive probability.

- **The Fundamental Theorem of Finite Markov Chains:**
  - A Markov Chain is good $\Leftrightarrow$ the corresponding graph is ergodic.
  - We will prove the $\Rightarrow$ part, by showing that non-ergodic Markov Chains are not good.

# Examples of "Bad" Markov Chains

- A Markov chains is not "good" if either:
  1. It does not converge to a unique distribution.
  2. It does converge to u.d., but some states in this distribution have zero probability.

# Bad case 1: Mutual Unreachabaility




Consider two initial distributions:

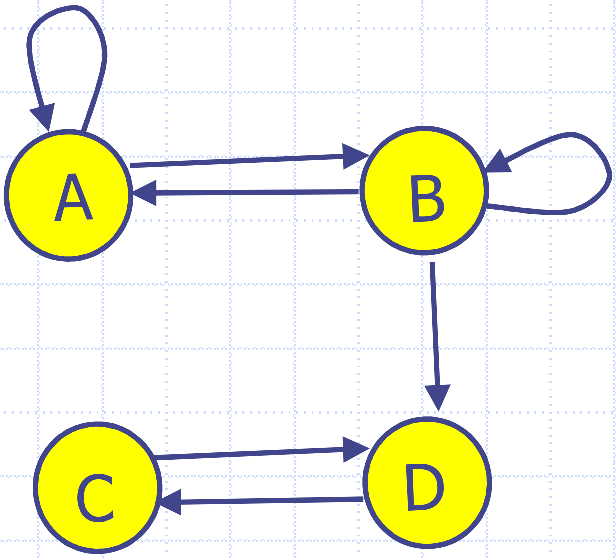
*a)* $p(X_1=A)=1$ ($p(X_1 = x)=0$ if $x \neq A$).

*b)* $p(X_1= C) = 1$

In case *a)*, the sequence will stay at A forever.
In case *b)*, it will stay in {C,D} for ever.

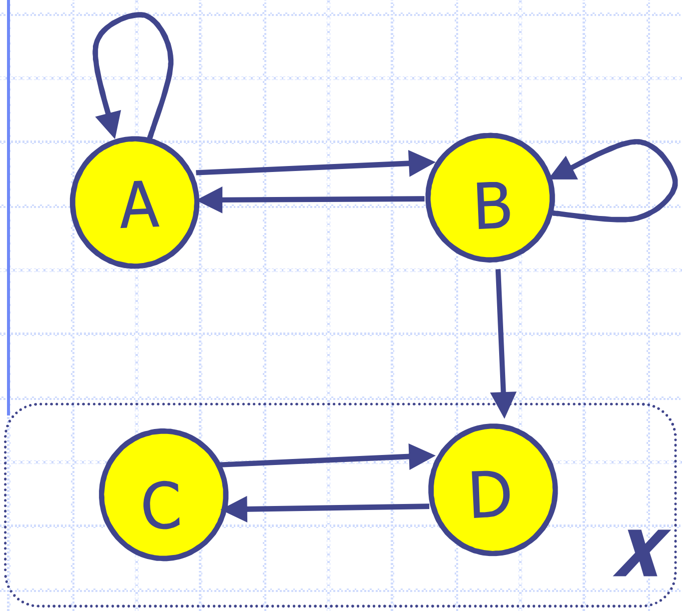
Fact 1: If G has two states which are unreachable from each other, then $\{X_i\}$ cannot converge to a distribution which is independent on the initial distribution.

# Bad case 2: Transient States




Once the process moves from **B** to **D**, it will never come back.

# Bad case 2: Transient States

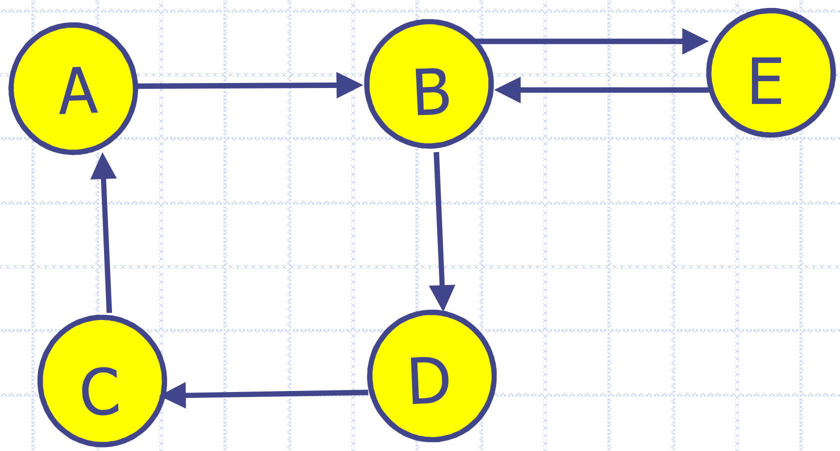



Fact 2: For each initial distribution, with probability 1 a transient state will be visited only a finite number of times.

Proof: Let A be a transient state, and let ***X*** be the set of states from which A is unreachable. It is enough to show that, starting from any state, with probability 1 a state in ***X*** is reached after a finite number of steps (Exercise: complete the proof)

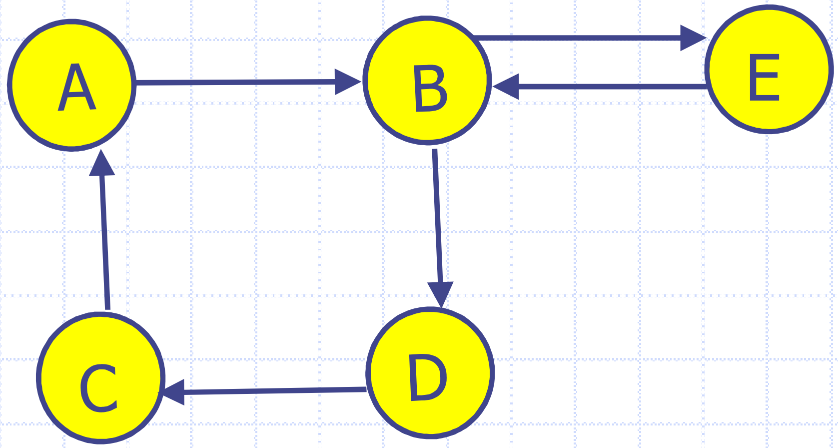
# Corollary: A good Markov Chain is irreducible

# Bad case 3: Periodic Markov Chains




Recall: A Markov Chain is ***periodic*** if all the states in it have a period $k$ >1. The above chain has period 2.
In the above chain, consider the initial distribution ***p***(B)=1.
Then states {B, C} are visited (with positive probability) only in odd steps, and states {A, D, E} are visited in only even steps.

# Bad case 3: Periodic States




Fact 3: In a periodic Markov Chain (of period $k > 1$) there are initial distributions under which the states are visited in a periodic manner. Under such initial distributions $\boldsymbol{X_i}$ does not converge as $\boldsymbol{i \rightarrow \infty}$.

Corollary: A good Markov Chain is not periodic

# The Fundamental Theorem of Finite Markov Chains:

- We have proved that non-ergodic Markov Chains are not good
- A proof of the other part (based on Perron-Frobenius theory) is beyond the scope of this course:

If a Markov Chain is ergodic, then

1. It has a unique stationary distribution vector $V > \underline{0}$, which is an Eigenvector of the transition matrix.
2. For any initial distribution, the distributions $\boldsymbol{X_i}$, as $\boldsymbol{i} \rightarrow \infty$, converges to $\boldsymbol{V}$.