Entering Answers in Pearson Example

- $eigenVector = \begin{bmatrix} 0.5144958 \\ 0.8574929 \end{bmatrix}$
- Note that $0.5144958/\ 0.8574929$ = 0.6
- You should enter your eigenvector as $\begin{bmatrix} 6 \\ 10 \end{bmatrix}$ or something similar with rational entries.

Knowledge is power.
Knowledge of eigenvalues …

can be worth about $500 billion.*

*:According to CNN Money.

Overview

Google Page Rank Algorithm

Markov Chain

Transition Matrix

Steady State Vector

Reading: 10-1 and 10-2 from Lay

Background

- Search engines such as Google have to do three basic things:
    1. Crawl the web and locate all web pages with public access.
    2. Find a subset of relevant webpage based on the keywords entered.
    3. Rate the importance of each page in the database.
- Our focus is on step 3: “how can one meaningfully define and quantify the importance of any given page?”

Background

What’s your favorite search engine? Bing, Google, Yahoo, …

When search engines went online in the 90’s google was clearly set apart from others.

Google was the most successful site in returning relevant sites on its first page.

With other search engines you often had to go through screen after screen of links to irrelevant web pages that just happened to match the search text.

Most search engines at the time used ***text based ranking*** **system.**

Text Based Ranking

A search engine kept an index of all web pages.

When a user typed in a query search, the engine browsed through its index and **counted** the occurrences of the key words in each web site.

The sites with the **highest** number of occurrences of the key words got displayed back to the user.

This approach is problematic!

Page rank

- The usefulness of a search engine depends on the *relevance* of the result set it gives back.
- There may be many web pages containing the search term; however some of them will be more relevant, popular, or respected than others.
- Part of the magic behind Google is its PageRank algorithm.
- It quantitatively rates the importance of each page on the web, allowing Google to rank the pages and thereby present to the user the more important pages first.

Larry Page

Basis of Google’s Algorithm

The basic idea was that a webpage was important if it had hyperlinks to it from other important pages.

The motion of a random surfer that moves from webpage to webpage by choosing which hyperlink to follow, can be modeled using *Markov chains*.

The pages that this random surfer visits more often has to be more important and thus more relevant if their content matches the terms of a search.

Page and his colleagues tried to find the *steady-state vector* for a particular Markov chain whose entries can be interpreted as the amount of time a random surfer will spend at each webpage.

The calculation of this steady state vector is the basis for google’s Page rank algorithm.

Markov Chain

Markov Chain

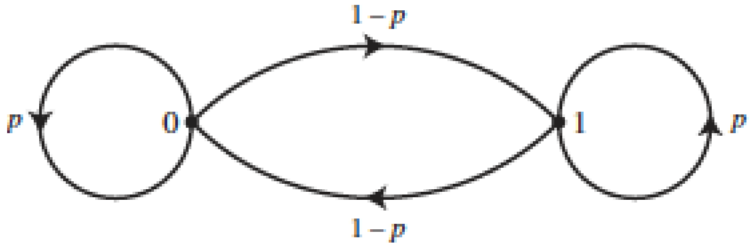
- It’s a mathematical model for movement between states.
- The process starts in one of these states and moves from state to state.

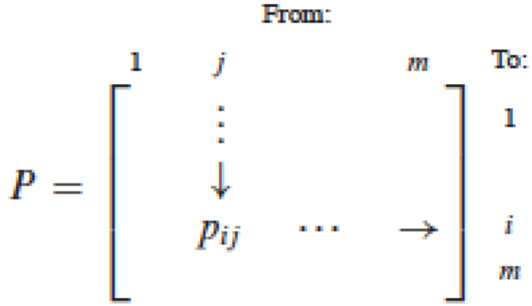
Transition Probability

- It’s the probability that the chain moves from state j to state i in one step.
- The state of the chain at any given state is not known, but the transition probability is.

Transition Matrix

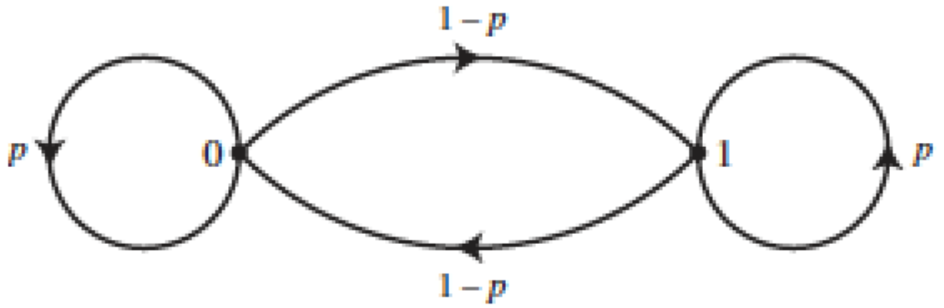
- It contains the transition probabilities.
- i-j[th] entry of the matrix denotes the probability of going from state j to state i.





Example

- Suppose that each bit of data is either 0 or a 1, and at each stage there is a probability p that the bit will pass through the state unchanged.
- Thus the probability is 1-p that the bit will be transposed.
- **Write the transition matrix.**

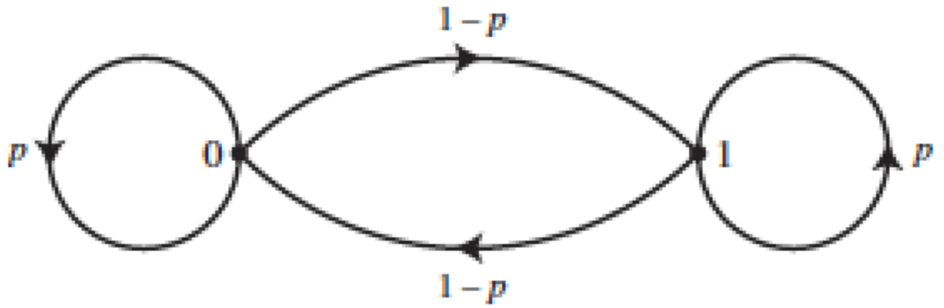


From:

$$\begin{array}{cc} & \begin{array}{cc} 0 & \quad 1 \end{array} \quad \text{To:} \\ P = & \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$

Example – cont.

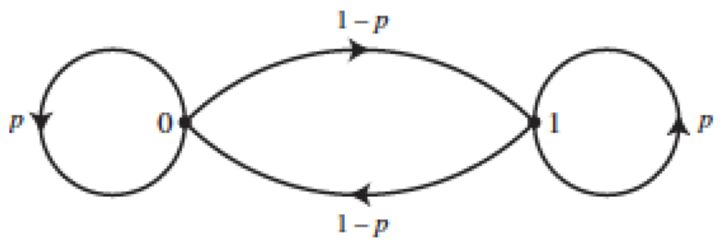
Suppose that p = 0.99. **Find a probability that the signal 0 will still be a 0 after a two-stage transmission process.**



Solution

- There are two possibilities to go to state 0 in 2 steps starting from state 0:
  1. $0 \;->\; 0 \;->\; 0 \Longrightarrow P = 0.99^2 = 0.9801$
  2. $0 \;->\; 1 \;->\; 0 \Longrightarrow P = (0.01)(0.01) = 0.0001$

  $0.9801 + 0.0001 = 0.9802$



- Note that we can also use the transition matrix and initial *state-vector* (prob. that chain is in each state initially) to find this value.
- $\begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix}$
- $\begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} 0.99 \\ 0.01 \end{bmatrix} = \begin{bmatrix} 0.99^2 + 0.01^2 \\ * \end{bmatrix}$

State-Vector X_n lists the probabilities that the chain is in each of the possible states after n steps.

For m possible states we will have:

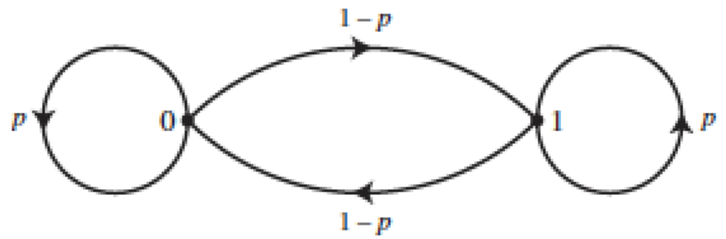
State vectors are probability vectors, therefore the sum of their entries must be one.

$$\mathbf{x}_n = \begin{bmatrix} a_1 \\ \vdots \\ a_j \\ \vdots \\ a_m \end{bmatrix} \longleftarrow \text{Probability that the chain is at state } j \text{ after } n \text{ steps}$$

State Vector

In the previous example $x_0=\begin{bmatrix}1\\0\end{bmatrix}$ means that ***initially*** probability that chain is in state zero is 1 and the probability that it's in state one is 0.

$x_1 = \begin{bmatrix}0.99\\0.01\end{bmatrix}$ means that ***after one step*** the probability that chain is still in state zero is .99 and the probability that it's in state one is 0.01.



State Vector - Example

Properties of Markov Chain

This state vectors for the chain are related by the equation:

$$\mathbf{x}_{n+1} = P\mathbf{x}_n$$

We can use the previous equation to show that:

$$\mathbf{x}_n = P^n\mathbf{x}_0$$

Markov chain is Memoryless:

Since the values in the vector X_{n+1} depend only on the transition matrix P and on X_n, the state of the chain before time n has zero effect on its state at the time n+1 and beyond.

Random Walk

Think of the states

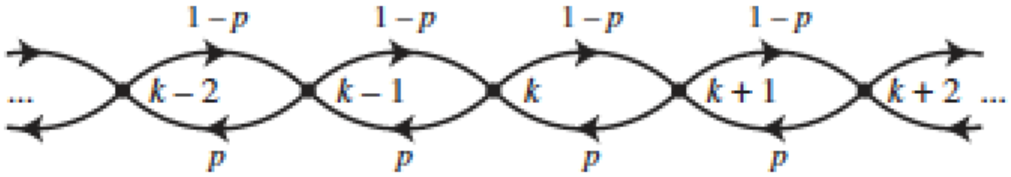
{ 1, 2,…,n} as lying on a line.

Place a molecule at a point that is not on the end of the line.

At each step the molecule moves left one unit with probability p and write one unit with probability 1- p.

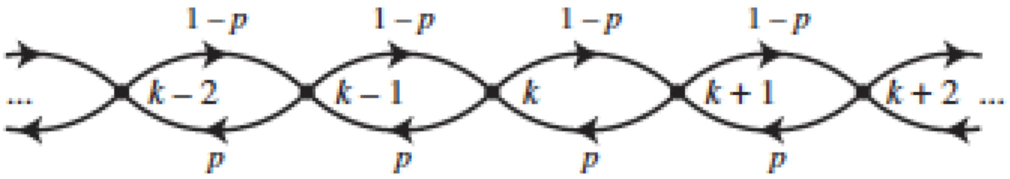
The molecule walks randomly along the line.

If p=0.5, the walk is called simple or unbiased, otherwise, the walk is said to be biased.



Absorbing Boundaries and Reflecting Boundaries

When at states 2,…,n-1, the molecule must move to either the left or the right , but it cannot do this at the end points 1 and n.



The molecule's possible movements at the end points 1 and n must be specified.

Absorbing Boundaries :

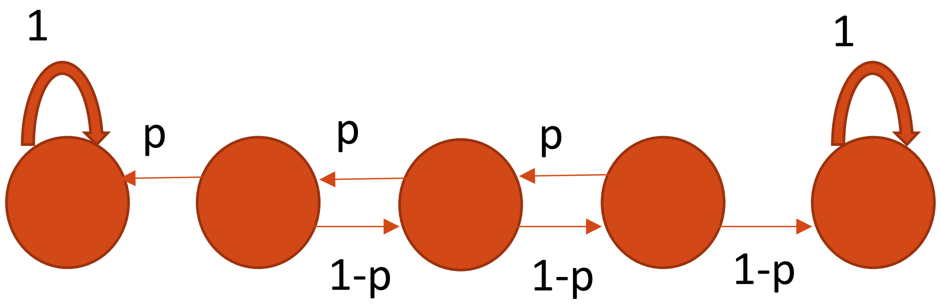
One possibility is to have the molecules stay at an endpoint forever once it reaches either end of the line. The endpoints 1 and n are called absorbing states.

Reflecting Boundaries:

Another possibility is to have the molecule bounce back one unit when an end point is reached. This is called a random walk with reflecting boundaries

Example

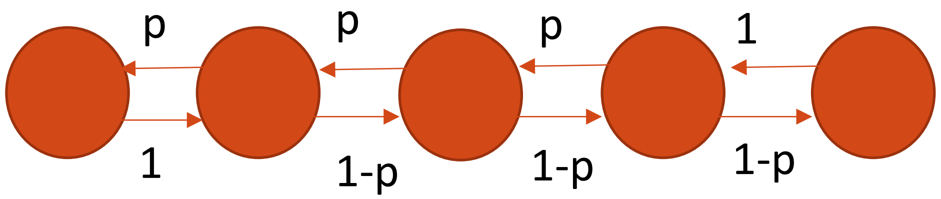
Draw the graph and write the transition matrix for a random walk on { 1, 2, 3, 4, 5} with absorbing boundaries.



$$P = \begin{array}{c} \\ \end{array}\begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 1 & p & 0 & 0 & 0 \\ 0 & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & 0 \\ 0 & 0 & 0 & 1-p & 1 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

Example

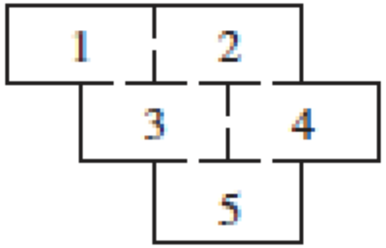
Draw the graph and write the transition matrix for a random walk on { 1, 2, 3, 4, 5} with reflecting boundaries.



$$P = \begin{array}{c} \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} \;\;1 & \;\;2 & \;\;3 & \;\;4 & \;\;5 \end{array} \\ \begin{bmatrix} 0 & p & 0 & 0 & 0 \\ 1 & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & 1 \\ 0 & 0 & 0 & 1-p & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Example

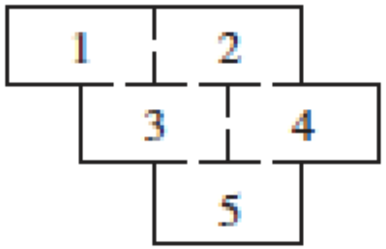
- Suppose a mouse runs through a 5-room maze in this figure.
- The mouse moves to a different room at each time step.
- When the mouse is in a particular room, it is equally likely to choose any of the doors out of the room.
- Find the transition matrix.



$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Example – cont.

Find the probability that a mouse starting in room 3 returns to that room in exactly 5 steps.

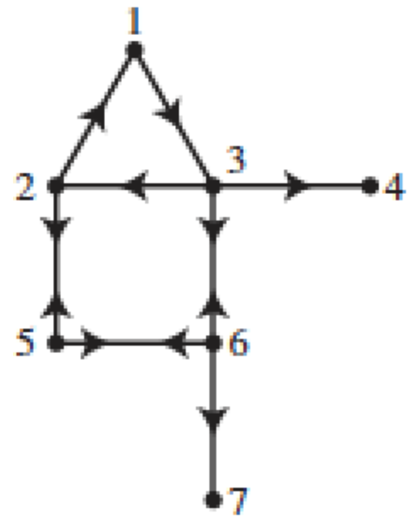


$$P = \begin{array}{c c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

$$\mathbf{x}_5 = P^5\mathbf{x}_0 = P^5\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} .1507 \\ .2143 \\ .2701 \\ .2143 \\ .1507 \end{bmatrix}$$

Example – cont.

- Consider a set of seven pages hyperlinked by the directed graph in this figure.
- Find the transition matrix.
- If the random surfers starts at page 5, find the probability that the surfer will be at page 3 after four clicks.
- Note that states 4 and 7 are absorbing states for this Markov chain.
- These are called ***dangling nodes*** and are quiet common on the web.



$$P = \begin{array}{c} \\ \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}$$

$$\mathbf{x}_4 = P^4\mathbf{x}_0 = \begin{bmatrix} .1319 \\ .0833 \\ .0880 \\ .1389 \\ .2199 \\ .0833 \\ .2546 \end{bmatrix}$$

Steady-State Vector

The most interesting aspect of a Markov chain is its long-term behavior of X_n.

In many cases the sequence of vectors $\{X_n\}$ converges to a vector. This vector is called the steady-state vector of the Markov chain.

Example

$$P = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix} \qquad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{x}_1 = P\mathbf{x}_0 = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} .5 \\ .3 \\ .2 \end{bmatrix}$$

These vectors seem to be approaching $\mathbf{q} = \begin{bmatrix} .3 \\ .6 \\ .1 \end{bmatrix}$.

$$\mathbf{x}_2 = P\mathbf{x}_1 = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix} \begin{bmatrix} .5 \\ .3 \\ .2 \end{bmatrix} = \begin{bmatrix} .37 \\ .45 \\ .18 \end{bmatrix}$$

$$\mathbf{x}_3 = P\mathbf{x}_2 = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix} \begin{bmatrix} .37 \\ .45 \\ .18 \end{bmatrix} = \begin{bmatrix} .329 \\ .525 \\ .146 \end{bmatrix}$$

$$\mathbf{x}_{12} = \begin{bmatrix} .30005 \\ .59985 \\ .10010 \end{bmatrix}, \quad \mathbf{x}_{13} = \begin{bmatrix} .30002 \\ .59993 \\ .10005 \end{bmatrix}, \quad \mathbf{x}_{14} = \begin{bmatrix} .30001 \\ .59996 \\ .10002 \end{bmatrix}, \quad \mathbf{x}_{15} = \begin{bmatrix} .30001 \\ .59998 \\ .10001 \end{bmatrix}$$

Example – cont.

Also note that P^n also converges to a matrix as n increases.

$$P^2 = \begin{bmatrix} .3700 & .2600 & .3300 \\ .4500 & .7000 & .4500 \\ .1800 & .0400 & .2200 \end{bmatrix} \qquad P^3 = \begin{bmatrix} .3290 & .2820 & .3210 \\ .5250 & .6500 & .5250 \\ .1460 & .0680 & .1540 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} .3133 & .2914 & .3117 \\ .5625 & .6250 & .5625 \\ .1242 & .0836 & .1258 \end{bmatrix} \qquad P^5 = \begin{bmatrix} .3064 & .2958 & .3061 \\ .5813 & .6125 & .5813 \\ .1123 & .0917 & .1127 \end{bmatrix}$$

$$P^{10} = \begin{bmatrix} .3002 & .2999 & .3002 \\ .5994 & .6004 & .5994 \\ .1004 & .0997 & .1004 \end{bmatrix} \qquad P^{15} = \begin{bmatrix} .3000 & .3000 & .3000 \\ .6000 & .6000 & .6000 \\ .1000 & .1000 & .1000 \end{bmatrix}$$

Definition: Stochastic Matrix and Steady-State Vector

- A is a **stochastic** matrix if all A's entries are nonnegative and the sum of the entries in each column is equal to 1.
- Any column-stochastic matrix has 1 as an eigenvalue.

If P is a stochastic matrix, then a **steady-state vector** (or **equilibrium vector** or **invariant probability vector**) for P is a probability vector $\mathbf{q}$ such that

$$P\mathbf{q} = \mathbf{q}$$

Example 1

- Find the steady-state vector for the mouse-in-the-maze example.

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

Solution

The steady-state vector may be computed by solving the system $P\mathbf{q} = \mathbf{q}$, which is equivalent to the homogeneous system $(P - I)\mathbf{q} = \mathbf{0}$. Row reduction yields

$$\begin{bmatrix} -1 & 1/3 & 1/4 & 0 & 0 & 0 \\ 1/2 & -1 & 1/4 & 1/3 & 0 & 0 \\ 1/2 & 1/3 & -1 & 1/3 & 1/2 & 0 \\ 0 & 1/3 & 1/4 & -1 & 1/2 & 0 \\ 0 & 0 & 1/4 & 1/3 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -3/2 & 0 \\ 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & -3/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$q_5 \begin{bmatrix} 1 \\ 3/2 \\ 2 \\ 3/2 \\ 1 \end{bmatrix}$$

$$\mathbf{q} = \frac{1}{7}\begin{bmatrix} 1 \\ 3/2 \\ 2 \\ 3/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/7 \\ 3/14 \\ 2/7 \\ 3/14 \\ 1/7 \end{bmatrix} \approx \begin{bmatrix} .142857 \\ .214286 \\ .285714 \\ .214286 \\ .142857 \end{bmatrix}$$

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/3 & 0 \\ 1/2 & 1/3 & 0 & 1/3 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 1/2 \\ 0 & 0 & 1/4 & 1/3 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

```
> (P = matrix(c(0,rep(1/2,2),rep(0,2),1/3,0,rep(1/3,2), 0, rep(1/4,2),0,rep(1/4,2),0,rep(1/3,2),0,1/3,rep(0,2),rep(1/2,2),0), nrow=5))
     [,1]      [,2] [,3]      [,4] [,5]
[1,]  0.0 0.3333333 0.25 0.0000000  0.0
[2,]  0.5 0.0000000 0.25 0.3333333  0.0
[3,]  0.5 0.3333333 0.00 0.3333333  0.5
[4,]  0.0 0.3333333 0.25 0.0000000  0.5
[5,]  0.0 0.0000000 0.25 0.3333333  0.0
> v=eigen(P)
> v$vectors
          [,1]          [,2]          [,3]          [,4]       [,5]
[1,] 0.3086067  3.400937e-01  4.082483e-01  5.460122e-01  0.2672612
[2,] 0.4629100 -6.199486e-01 -9.791450e-16  4.493002e-01 -0.5345225
[3,] 0.6172134  2.686535e-15 -8.164966e-01  1.207994e-16  0.5345225
[4,] 0.4629100  6.199486e-01  1.674921e-15 -4.493002e-01 -0.5345225
[5,] 0.3086067 -3.400937e-01  4.082483e-01 -5.460122e-01  0.2672612
```

```
> v$vectors[,1]/sum(v$vectors[,1])
[1] 0.1428571 0.2142857 0.2857143 0.2142857 0.1428571
```

Interpretations of the steady-state vector

- After many moves the probability that the mouse will be in room one at a given time is approximately 1/7 no matter where the mouse began its journey.
- The mouse is expected to be in room one, 1/7th or about 14.3% of the time.

$$\mathbf{q} = \frac{1}{7}\begin{bmatrix} 1 \\ 3/2 \\ 2 \\ 3/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/7 \\ 3/14 \\ 2/7 \\ 3/14 \\ 1/7 \end{bmatrix} \approx \begin{bmatrix} .142857 \\ .214286 \\ .285714 \\ .214286 \\ .142857 \end{bmatrix}$$

Again notice that taking high powers of the transition matrix P gives matrices whose columns are converging to $\mathbf{q}$; for example,

$$P^{10} = \begin{bmatrix} .144169 & .141561 & .142613 & .144153 & .142034 \\ .212342 & .216649 & .214286 & .211922 & .216230 \\ .285226 & .285714 & .286203 & .285714 & .285226 \\ .216230 & .211922 & .214286 & .216649 & .212342 \\ .142034 & .144153 & .142613 & .141561 & .144169 \end{bmatrix}$$

Some difficulties…

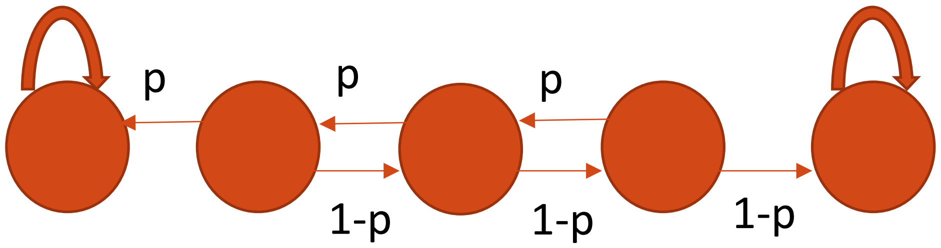
Every stochastic matrix will have a steady state vector, but in some cases a steady-state vectors cannot be interpreted as vectors of long run probabilities or of occupation times.

Example 2: Absorbing Boundaries

- Consider an unbiased random walk on { 1, 2, 3, 4, 5) with absorbing boundaries and the following transition matrix.

$$P = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

- Find P^{20} and P^{30}.
- Do columns of P^n seem to converge to a unique vector q?



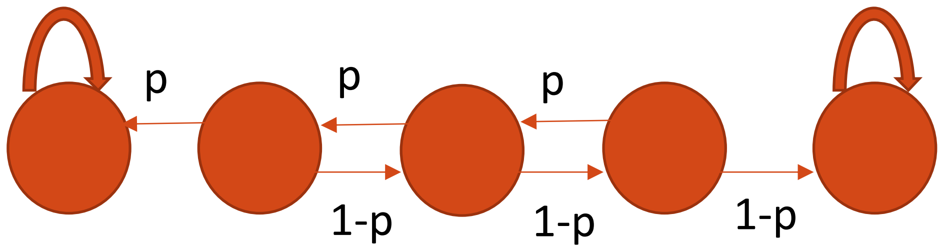
Solution

```
> library(expm)
> (P = matrix(c(1,rep(0,4),1/2,0,1/2,rep(0,3),1/2,0,1/2, rep(0,3),1/2,0,1/2, rep(0,4),1),nrow=5))
     [,1] [,2] [,3] [,4] [,5]
[1,]    1  0.5  0.0  0.0    0
[2,]    0  0.0  0.5  0.0    0
[3,]    0  0.5  0.0  0.5    0
[4,]    0  0.0  0.5  0.0    0
[5,]    0  0.0  0.0  0.5    1
```

```
> P%^%20
     [,1]         [,2]         [,3]         [,4] [,5]
[1,]    1 0.7495117188 0.4995117188 0.2495117188    0
[2,]    0 0.0004882812 0.0000000000 0.0004882812    0
[3,]    0 0.0000000000 0.0009765625 0.0000000000    0
[4,]    0 0.0004882812 0.0000000000 0.0004882812    0
[5,]    0 0.2495117188 0.4995117188 0.7495117188    1
```

Notice that only two long-term possibilities exist for this chain: it must end up in state 1 or state 5. Thus the probability that the chain is in state 2, 3, or 4 becomes smaller and smaller as n increases, as P^n illustrates:

```
> P%^%30
     [,1]         [,2]         [,3]         [,4] [,5]
[1,]    1 7.499847e-01 4.999847e-01 2.499847e-01    0
[2,]    0 1.525879e-05 0.000000e+00 1.525879e-05    0
[3,]    0 0.000000e+00 3.051758e-05 0.000000e+00    0
[4,]    0 1.525879e-05 0.000000e+00 1.525879e-05    0
[5,]    0 2.499847e-01 4.999847e-01 7.499847e-01    1
```



Solution

It seems that P^n converges to the matrix

$$\begin{bmatrix} 1 & .75 & .5 & .25 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & .25 & .5 & .75 & 1 \end{bmatrix}$$

as n increases. But the columns of this matrix are not equal; the probability of ending up either at 1 or at 5 depends on where the chain begins.

Solution

$0 \le q \le 1$ the vector

$$\begin{bmatrix} q \\ 0 \\ 0 \\ 0 \\ 1-q \end{bmatrix}$$

is a steady-state vector for P. This matrix then has an infinite number of possible steady-state vectors, which shows in another way that $\mathbf{x}_n$ cannot be expected to have convergent behavior which does not depend on $\mathbf{x}_0$. ■

```
> rref(P-diag(5))
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    1    0    0    0
[2,]    0    0    1    0    0
[3,]    0    0    0    1    0
[4,]    0    0    0    0    0
[5,]    0    0    0    0    0
```

From above we'll have:

$$x_2 = 0$$
$$x_3 = 0$$
$$x_4 = 0$$

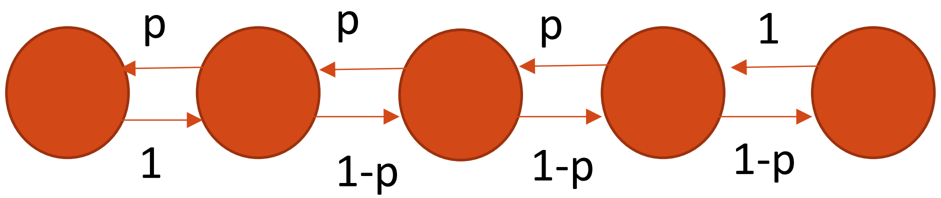
Note that x_1 and x_5 are free parameters,
But since the matrix is stochastic
$x_1 + x_5$ should be equal to 1.

Example 3: Reflecting Boundaries

- Consider an unbiased random walk on { 1, 2, 3, 4, 5) with reflecting boundaries and the following transition matrix.

$$P = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array}$$

- Find P^{20} and P^{21}.
- Does P^n seem to converge to a unique matrix?



If the chain $\mathbf{x}_n$ starts at state 1, notice that it can return to 1 only when n is even, while the chain can be at state 2 only when n is odd.

Solution

P^n cannot converge to a unique matrix since P^n looks very different depending on whether n is even or odd, as shown:

$$P^{20} = \begin{bmatrix} .2505 & 0 & .2500 & 0 & .2495 \\ 0 & .5005 & 0 & .4995 & 0 \\ .5000 & 0 & .5000 & 0 & .5000 \\ 0 & .4995 & 0 & .5005 & 0 \\ .2495 & 0 & .2500 & 0 & .2505 \end{bmatrix}$$

$$P^{21} = \begin{bmatrix} 0 & .2502 & 0 & .2498 & 0 \\ .5005 & 0 & .5000 & 0 & .4995 \\ 0 & .5000 & 0 & .5000 & 0 \\ .4995 & 0 & .5000 & 0 & .5005 \\ 0 & .2498 & 0 & .2502 & 0 \end{bmatrix}$$

Even though P^n does not converge to a unique matrix, P does have a steady-state vector.

```
> (P = matrix(c(0,1,rep(0,3),1/2,0,1/2, rep(0,3),1/2,0,1/2,rep(0,3),1/2,0,1/2,rep(0,3),1,0),nrow=5))
     [,1] [,2] [,3] [,4] [,5]
[1,]    0  0.5  0.0  0.0    0
[2,]    1  0.0  0.5  0.0    0
[3,]    0  0.5  0.0  0.5    0
[4,]    0  0.0  0.5  0.0    1
[5,]    0  0.0  0.0  0.5    0
> P%^%20
          [,1]      [,2] [,3]      [,4]      [,5]
[1,] 0.2504883 0.0000000 0.25 0.0000000 0.2495117
[2,] 0.0000000 0.5004883 0.00 0.4995117 0.0000000
[3,] 0.5000000 0.0000000 0.50 0.0000000 0.5000000
[4,] 0.0000000 0.4995117 0.00 0.5004883 0.0000000
[5,] 0.2495117 0.0000000 0.25 0.0000000 0.2504883
> P%^%21
          [,1]      [,2] [,3]      [,4]      [,5]
[1,] 0.0000000 0.2502441  0.0 0.2497559 0.0000000
[2,] 0.5004883 0.0000000  0.5 0.0000000 0.4995117
[3,] 0.0000000 0.5000000  0.0 0.5000000 0.0000000
[4,] 0.4995117 0.0000000  0.5 0.0000000 0.5004883
[5,] 0.0000000 0.2497559  0.0 0.2502441 0.0000000
> v=eigen(P)
> v$vectors
          [,1]       [,2]          [,3]          [,4]          [,5]
[1,] 0.2672612  0.2672612  4.082483e-01 -4.082483e-01  4.082483e-01
[2,] 0.5345225 -0.5345225 -5.773503e-01 -5.773503e-01  5.417334e-17
[3,] 0.5345225  0.5345225  9.488593e-17  6.644423e-16 -8.164966e-01
[4,] 0.5345225 -0.5345225  5.773503e-01  5.773503e-01  1.218900e-16
[5,] 0.2672612  0.2672612 -4.082483e-01  4.082483e-01  4.082483e-01
> v$vectors[,1]/sum(v$vectors[,1])
[1] 0.125 0.250 0.250 0.250 0.125
```

Solution Using rref

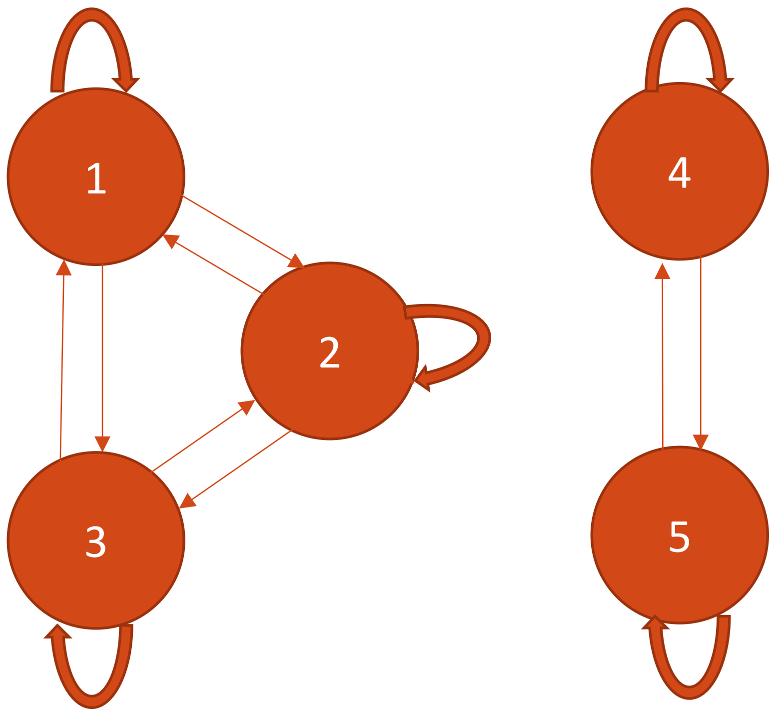
```
> rref(P-diag(5))
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0   -1
[2,]    0    1    0    0   -2
[3,]    0    0    1    0   -2
[4,]    0    0    0    1   -2
[5,]    0    0    0    0    0
```

- $x_1 - x_5 = 0$
- $x_2 - 2x_5 = 0$
- $x_3 - 2x_5 = 0$
- $x_4 - 2x_5 = 0$
- $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_5 \begin{bmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/8 \\ 2/8 \\ 2/8 \\ 2/8 \\ 1/8 \end{bmatrix}$ which is the same result as before:

Example 4: Disconnected Nodes

- Find the steady-state vector for the following transition matrix.

$$
P = \begin{array}{c c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\
\begin{bmatrix}
1/4 & 1/3 & 1/2 & 0 & 0 \\
1/4 & 1/3 & 1/4 & 0 & 0 \\
1/2 & 1/3 & 1/4 & 0 & 0 \\
0 & 0 & 0 & 1/3 & 3/4 \\
0 & 0 & 0 & 2/3 & 1/4
\end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\end{array}
$$

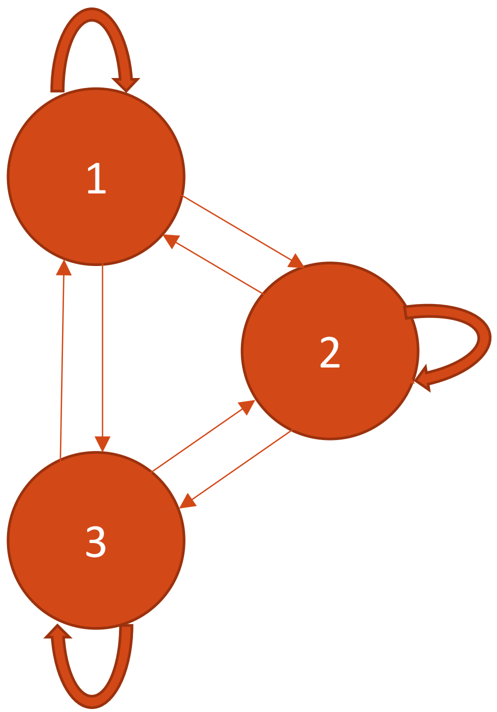


Solution

If this Markov chain begins at state 1, 2, or 3, then it must always be at one of those states. Likewise if the chain starts at state 4 or 5, then it must always be at one of those states. The chain splits into two separate chains, each with its own steady-state vector. In this case P^n converges to a matrix whose columns are not equal. The vectors

$$\begin{bmatrix} 4/11 \\ 3/11 \\ 4/11 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 9/17 \\ 8/17 \end{bmatrix}$$

both satisfy the definition of steady-state vector



The first vector gives the limiting probabilities if the chain starts at state 1, 2, or 3, and the second does the same for states 4 and 5. ■

Solution in R

```
> P = matrix(c(1/4,1/4,1/2,rep(0,2), rep(1/3,3),rep(0,2),1/2,1/4,1/4,rep(0,2),rep(0,3),1/3,2/3, rep(0,3),3/4,1/4),nrow = 5)
> (v=eigen(P))
eigen() decomposition
$values
[1]  1.00000000  1.00000000 -0.41666667 -0.25000000  0.08333333

$vectors
          [,1]       [,2]       [,3]          [,4]       [,5]
[1,] 0.0000000 -0.6246950  0.0000000 -7.071068e-01  0.4082483
[2,] 0.0000000 -0.4685213  0.0000000  4.666006e-17 -0.8164966
[3,] 0.0000000 -0.6246950  0.0000000  7.071068e-01  0.4082483
[4,] 0.7474093  0.0000000 -0.7071068  0.000000e+00  0.0000000
[5,] 0.6643638  0.0000000  0.7071068  0.000000e+00  0.0000000

> sweep(v$vectors, MARGIN = 2, colSums(v$vectors), '/')
          [,1]      [,2]          [,3]          [,4]          [,5]
[1,] 0.0000000 0.3636364  0.000000e+00 -1.514961e+16 -1.470869e+15
[2,] 0.0000000 0.2727273  0.000000e+00  9.996821e-01  2.941739e+15
[3,] 0.0000000 0.3636364  0.000000e+00  1.514961e+16 -1.470869e+15
[4,] 0.5294118 0.0000000  3.184526e+15  0.000000e+00  0.000000e+00
[5,] 0.4705882 0.0000000 -3.184526e+15  0.000000e+00  0.000000e+00
```

Observation from previous 3 examples

1

In the examples with absorbing boundaries, and disconnected nodes, **the steady-state vector is not unique.**

2

In all 3 previous examples the matrix $\mathbf{P^N}$ **does not converge** to a matrix with equal columns as N increases.

3

In all these examples, **every matrix** of the form P^k has some zero entries.

4

If this wasn't the case, i.e. some $\mathbf{P^k}$ **contained only positive entri**es then for some k , it was possible to move from any state to any other in exactly k steps.

Regular Matrix

A stochastic matrix P is **regular** if some power P^k contains only strictly positive entries.

A Markov chain with a regular transition matrix has the following properties that we are interested in:

a. There is a stochastic matrix Π such that $\lim_{n\to\infty} P^n = \Pi$.

b. Each column of Π is the same probability vector $\mathbf{q}$.

c. For any initial probability vector $\mathbf{x}_0$, $\lim_{n\to\infty} P^n\mathbf{x}_0 = \mathbf{q}$.

d. The vector $\mathbf{q}$ is the unique probability vector which is an eigenvector of P associated with the eigenvalue 1.

e. All eigenvalues λ of P other than 1 have $|\lambda| < 1$.

Regular matrices, Markov chain, and Google Page Rank

We mentioned earlier the World Wide Web can be modeled as a directed graph, with vertices representing the webpages and the arrows representing the links between webpages.

If the matrix P for this Markov chain was regular then there exist a steady-state vector q whose entries would tell *what fraction of the random surfer's time was spent at each webpage.*

The idea behind google algorithm is that important pages have links coming from other important pages. Therefore, a random surfer will spend more time at more important pages and less time at less important pages.

This means that the amount of time to spend at each page is just the occupation time for each state in the Markov chain.

The importance of a webpage may be measured by the relative size of the corresponding entry in the steady-state vector **q** for an appropriately chosen Markov chain.

ADJUSTMENT 1: If the surfer reaches a dangling node, the surfer will pick any page in the Web with equal probability and will move to that page. In terms of the transition matrix P, if state j is an absorbing state, replace column j of P with the vector

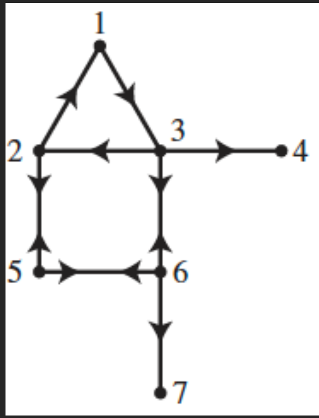
$$\begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

where n is the number of rows (and columns) in P.

Modifications to Address Dangling Nodes

Example

Write the transition matrix for the following 7-page Web.



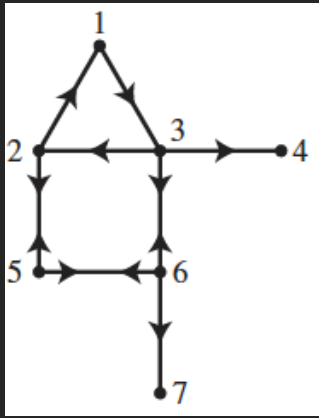
$$P_* = \begin{array}{c} \begin{matrix} 1 & \;\;2 & \;\;3 & \;\;4 & \;\;5 & \;\;6 & \;\;7 \end{matrix} \\ \begin{bmatrix} 0 & 1/2 & 0 & 1/7 & 0 & 0 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 1 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 0 & 0 & 1/7 \\ 0 & 1/2 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 0 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \end{bmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}$$

- In order to overcome disconnected nodes, fix a positive constant p between 0 and 1 (typically 0.15), called damping factor.
- Define the Google Matrix of the graph by PageRank matrix where
- $\boldsymbol{G = (1-p)P + pK} \quad where \quad K = \frac{1}{k}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$
- The pK term indicates that with probability p, the surfer will pick any page in the web with equal probability and will move to that page.
- Matrix G is guaranteed to be regular.

Adjustment 2: Modifications to Address Disconnected Nodes

Example

- Write ***Google Matrix*** for the following 7-page Web.
- Rank the webpages from highest to lowest.



$$G = .85\begin{bmatrix} 0 & 1/2 & 0 & 1/7 & 0 & 0 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 1 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 0 & 0 & 1/7 \\ 0 & 1/2 & 0 & 1/7 & 0 & 1/3 & 1/7 \\ 0 & 0 & 1/3 & 1/7 & 1/2 & 0 & 1/7 \\ 0 & 0 & 0 & 1/7 & 0 & 1/3 & 1/7 \end{bmatrix}$$

$$+ .15\begin{bmatrix} 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \end{bmatrix}$$

$$= \begin{bmatrix} .021429 & .446429 & .021429 & .142857 & .021429 & .021429 & .142857 \\ .021429 & .021429 & .304762 & .142857 & .446429 & .021429 & .142857 \\ .871429 & .021429 & .021429 & .142857 & .021429 & .304762 & .142857 \\ .021429 & .021429 & .304762 & .142857 & .021429 & .021429 & .142857 \\ .021429 & .446429 & .021429 & .142857 & .021429 & .304762 & .142857 \\ .021429 & .021429 & .304762 & .142857 & .446429 & .021429 & .142857 \\ .021429 & .021429 & .021429 & .142857 & .021429 & .304762 & .142857 \end{bmatrix}$$

It is now possible to find the steady-state vector by the methods of this section:

$$\mathbf{q} = \begin{bmatrix} .116293 \\ .168567 \\ .191263 \\ .098844 \\ .164054 \\ .168567 \\ .092413 \end{bmatrix}$$

so the most important page according to PageRank is page 3, which has the largest entry in $\mathbf{q}$. The complete ranking is 3, 2 and 6, 5, 1, 4, and 7.

Solution