A Look at Markov Chains and Their Use in Google

Rebecca Atherton

Iowa State University

MSM Creative Component

Summer 2005

Leslie Hogben, Major Professor

Heather Thompson, Committee Member

Loren Zachary, Committee Member

Section 1

Introduction to Markov Chains

The topic of matrices is an interesting one to teach to Algebra II and Pre-Calculus students. Students are often able to master the algorithms associated with matrices, but they are sometimes left wondering where matrices would be used in the "real world". Matrices have many "real world" applications; one particularly interesting application is the Markov chain. In this section we give a general introduction to Markov chains and their properties by examining examples. In Section 2 we discuss how the internet search engine Google uses Markov chains to rank the pages discovered in a user's search. Section 3 presents a careful treatment of the mathematics of Markov chains.

Markov chains were named after their inventor, A. A. Markov, a Russian Mathematician who worked in the early 1900's. Simply put, a Markov chain uses a matrix and a vector (column matrix) to model and predict the behavior of a system that moves from one state to another state in a way that depends only on the current state.

To appreciate the power of Markov chains, let us begin with an example given by the NCTM in New Topics for Secondary School Mathematics: Matrices. Suppose a taxi company has divided the city into three regions: Northside, Downtown, and Southside. The company has been keeping track of pickups and deliveries and has found that of the fares picked up Downtown, 10% are dropped off in Northside, 40% stay Downtown, and 50% go to Southside. Of the fares that originate in Northside, 50% stay in Northside, 20% are dropped off in Downtown, and 30% are dropped off in Southside. Of those fares picked up in Southside, 30% end in Northside, 30% are delivered to Downtown, and 40% stay in the Southside region. The taxi company would like to know what the distribution of their taxis would be over a certain amount of time [20].

The picture below depicts the situation. (D represents the Downtown region, S is for the Southside, and N is for the Northside region.) In this picture, the arrows indicate the probability of moving from one state to another. "A state is the condition or location of an object in the system at a particular time. Therefore, our diagram is called a *state diagram*." [5] Figure 1 can also be thought of as a *directed graph* or *digraph*.

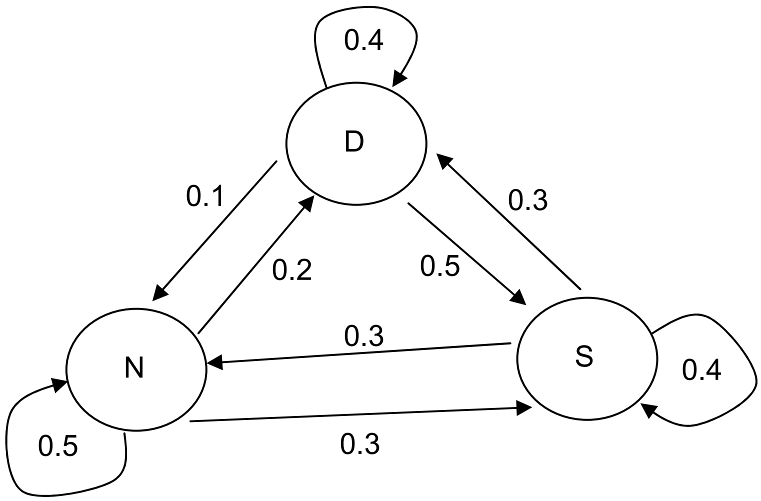


Figure 1: state diagram for taxi problem [20]

Suppose that a taxi begins its shift in the Northside Region. What is the probability that it will still be in the Northside Region after two fares? There are three ways that a taxi can start in the Northside Region and end up in the Northside Region. It can go from Northside to Southside and then back to Northside. It can go from Northside to Downtown and back to Northside. Or it can pick up in Northside, and stay there. If we use the notation $P(NN_2)$ to mean the probability of beginning in the Northside and ending in the Northside after two fares, and P(NS) to indicate the probability of traveling from Northside to Southside, P(ND) to mean the probability of traveling from Northside to Downtown, etc., we can write

$$\begin{aligned} P(NN_2) &= P(NN)P(NN) + P(ND)P(DN) + P(NS)P(SN) \\ &= (0.5)(0.5) + (0.2)(0.1) + (0.3)(0.3) \\ &= 0.25 + 0.02 + 0.09 = 0.36 \end{aligned}$$

Likewise, we can determine the probability of a taxicab beginning the night in the Northside region and ending up in the Downtown region after two fares by

$$\begin{aligned}P(ND_2) &= P(NN)P(ND) + P(ND)P(DD) + P(NS)P(SD)\\ &= (0.5)(0.2) + (0.2)(0.4) + (0.3)(0.3)\\ &= 0.1 + 0.08 + 0.09 = 0.27\end{aligned}$$

And the probability of a taxi beginning in Northside and ending up in the Southside region after two fares by

$$\begin{aligned}P(NS_2) &= P(NN)P(NS) + P(ND)P(DS) + P(NS)P(SS)\\ &= (0.5)(0.3) + (0.2)(0.5) + (0.3)(0.4)\\ &= 0.15 + 0.1 + 0.12 = 0.37\end{aligned}$$

Note that $P(NN_2) + P(ND_2) + P(NS_2) = 1$ since the taxi would have to be in one of the three locations after 2 fares. We could perform similar calculations to determine where the taxi would be after two fares with a different starting location.

Calculating the probabilities of where the taxi will be located after 2 fares was not particularly tedious. Imagine though, the calculations involved if we wanted to consider where the taxi would be after 5 or 10 or more fares! There must be an easier way.

Suppose we placed the state probability information into a matrix as follows:

$$\begin{array}{c} \\ N \\ D \\ S \end{array}\begin{array}{c} \begin{array}{ccc} N & D & S \end{array} \\ \begin{bmatrix} .5 & .1 & .3 \\ .2 & .4 & .3 \\ .3 & .5 & .4 \end{bmatrix} \end{array} = T$$

where t_{ij} is the probability of moving from state j to state i. In other words, j represents the starting location and i represents the ending location. T is often called a *transition matrix* because "it expresses the probability of movement (transition) from one state to another" [5]. Notice that the sum of each of the columns in the transition matrix is equal to one. Recall that this must be true since a taxi must be in one of the three locations after one fare.

Take another look at how we calculated the probability of starting in the Northside region and finishing in the Northside region after 2 fares:

$$\begin{aligned}P(NN_2) &= P(NN)P(NN) + P(ND)P(DN) + P(NS)P(SN)\\ &= (0.5)(0.5) + (0.2)(0.1) + (0.3)(0.3)\\ &= 0.25 + 0.02 + 0.09 = 0.36\end{aligned}$$

The probabilities of starting in Northside and moving to another location after the first fare, P(NN), P(ND), and P(NS), are given in the first column of the matrix *T*. Notice too, that the probabilities of moving from the location after the first fare, back to the Northside are given in the first row of matrix *T*. Thus, if matrix multiplication is performed, looking only at multiplying the information from row one by the information in column one, we can see the following:

$$\begin{bmatrix} .5 & .1 & .3 \\ . & . & . \\ . & . & . \end{bmatrix} \cdot \begin{bmatrix} .5 & . & . \\ .2 & . & . \\ .3 & . & . \end{bmatrix} = \begin{bmatrix} (.5)(.5)+(.1)(.2)+(.3)(.3) & . & . \\ . & . & . \\ . & . & . \end{bmatrix} = \begin{bmatrix} .36 & . & . \\ . & . & . \\ . & . & . \end{bmatrix}$$

In a similar fashion we can use matrix multiplication to calculate the probability of starting in Northside and ending in Downtown after two fares:

$$\begin{bmatrix} . & . & . \\ .2 & .4 & .3 \\ . & . & . \end{bmatrix} \cdot \begin{bmatrix} .5 & . & . \\ .2 & . & . \\ .3 & . & . \end{bmatrix} = \begin{bmatrix} . & . & . \\ (.2)(.5)+(.4)(.2)+(.3)(.3) & . & . \\ . & . & . \end{bmatrix} = \begin{bmatrix} . & . & . \\ .27 & . & . \\ . & . & . \end{bmatrix}$$

and starting in Northside and ending in the Southside after two fares:

$$\begin{bmatrix} . & . & . \\ . & . & . \\ .3 & .5 & .4 \end{bmatrix} \cdot \begin{bmatrix} .5 & . & . \\ .2 & . & . \\ .3 & . & . \end{bmatrix} = \begin{bmatrix} . & . & . \\ . & . & . \\ (.3)(.5)+(.5)(.2)+(.4)(.3) & . & . \end{bmatrix} = \begin{bmatrix} . & . & . \\ . & . & . \\ .37 & . & . \end{bmatrix}.$$

In fact, we can create a new matrix that will tell us the probabilities of ending up in any of the three locations after two fares, by calculating the following:

$$T^2 = TT = \begin{bmatrix} .5 & .1 & .3 \\ .2 & .4 & .3 \\ .3 & .5 & .4 \end{bmatrix} \begin{bmatrix} .5 & .1 & .3 \\ .2 & .4 & .3 \\ .3 & .5 & .4 \end{bmatrix} = \begin{bmatrix} .36 & .24 & .3 \\ .27 & .33 & .3 \\ .37 & .43 & .4 \end{bmatrix}$$

From this answer we can see that the probability of the taxi beginning in the Southside and ending up in the Downtown region after two fares is 0.3 or 30%.

Expanding upon this idea of matrix multiplication allows us to create a matrix that tells us the probability of the taxi being in the three locations after three fares.

$$T^3 = T^2T = \begin{bmatrix} .36 & .24 & .3 \\ .27 & .33 & .3 \\ .37 & .43 & .4 \end{bmatrix} \begin{bmatrix} .5 & .1 & .3 \\ .2 & .4 & .3 \\ .3 & .5 & .4 \end{bmatrix} = \begin{bmatrix} .318 & .282 & .3 \\ .291 & .309 & .3 \\ .391 & .409 & .4 \end{bmatrix}$$

The probability of the taxi being in the three locations after four fares would then be

$$T^4 = \begin{bmatrix} .3054 & .2946 & .3 \\ .2973 & .3027 & .3 \\ .3973 & .4027 & .4 \end{bmatrix}$$

After 5 fares,

$$T^5 = \begin{bmatrix} .30162 & .29838 & .3 \\ .29919 & .30081 & .3 \\ .39919 & .40081 & .4 \end{bmatrix}.$$

After 10 fares,

$$T^{10} = \begin{bmatrix} .300003 & .299996 & .3 \\ .299998 & .300002 & .3 \\ .399998 & .400002 & .4 \end{bmatrix}.$$

What is happening to our transition matrix as the number of fares increases? It appears as though the matrix is beginning to converge towards particular numbers. This tells us that after a large number of fares, it no longer matters which region the taxi started in. We will discuss more about the convergence of a transition matrix later.

Suppose we are told the initial distributions of the taxis in town: 20% of the taxis begin the shift in the Northside region, 50% start in Downtown, and 30% begin in the Southside. This initial distribution information can be written as an *initial probability vector*, $\mathbf{q} = \begin{bmatrix} .20 \\ .50 \\ .30 \end{bmatrix}$. Using the initial distribution information, we can calculate the percentage of cars in each region after a given number of fares. For example, suppose we wanted to know how many taxis are in the Downtown region after one fare. We can perform the following calculation:

$0.20P(ND) + 0.50P(DD) + 0.30P(SD) = 0.20(0.2) + 0.50(0.4) + 0.30(0.3) = 0.33$. This means that after one fare, 33% of the taxis are in the Downtown region. Notice too that this sum is the product of the second row of our transition matrix and **q**: $\begin{bmatrix} . & . & . \\ .2 & .4 & .3 \\ . & . & . \end{bmatrix} . \begin{bmatrix} .2 \\ .5 \\ .3 \end{bmatrix} = \begin{bmatrix} . \\ .33 \\ . \end{bmatrix}$. We could also find the percentage of taxis located in the Northside and Southside regions by multiplying the first row of *T* by **q** and the third row of *T* by **q**, respectively.

Thus, the entire distribution of taxis after one fare, $\mathbf{q^1}$, can be determined by multiplying our transition matrix by the initial probability vector, $\mathbf{q^1} = T\mathbf{q} = \begin{bmatrix} .5 & .1 & .3 \\ .2 & .4 & .3 \\ .3 & .5 & .4 \end{bmatrix} \begin{bmatrix} .2 \\ .5 \\ .3 \end{bmatrix} = \begin{bmatrix} .24 \\ .33 \\ .43 \end{bmatrix}$. We can see that after one fare, the probability of a taxi being in the Northside region is 24%, the probability of being Downtown is 33%, and the probability of ending in the Southside region is 43%.

Now suppose we want to know where the taxis are after two fares, $\mathbf{q^2}$,

$$\mathbf{q^2} = T\mathbf{q^1} = TT\mathbf{q} = T^2\mathbf{q} = \begin{bmatrix} .36 & .24 & .3 \\ .27 & .33 & .3 \\ .37 & .43 & .4 \end{bmatrix} \begin{bmatrix} .2 \\ .5 \\ .3 \end{bmatrix} = \begin{bmatrix} .282 \\ .309 \\ .409 \end{bmatrix}$$

or after three fares, $\mathbf{q^3}$, $\mathbf{q^3} = T\mathbf{q^2} = TT^2\mathbf{q} = T^3\mathbf{q}$. Using a similar strategy, we can find the distribution after four fares by multiplying T^4 and $\mathbf{q}$:

$$\mathbf{q^4} = T^4\mathbf{q} = \begin{bmatrix} .3054 & .2946 & .3 \\ .2973 & .3027 & .3 \\ .3973 & .4027 & .4 \end{bmatrix} \begin{bmatrix} .2 \\ .5 \\ .3 \end{bmatrix} = \begin{bmatrix} .29838 \\ .30081 \\ .40081 \end{bmatrix}.$$

We can, in fact, determine the distribution after any *n* fares by using the equation $\mathbf{q}^n = T^n\mathbf{q}$, where $\mathbf{q}^n$ is commonly called the *state vector* since it represents the system's state after *n* steps.

In NCTM's taxi problem, the taxi company might be interested in the long-term distribution of the taxis. Using our equation $\mathbf{q^n} = T^n\mathbf{q}$, we can begin iterating:

After 5 fares: $\mathbf{q}^{\mathbf{5}} = T^{\mathbf{5}}\mathbf{q} = \begin{bmatrix} .299514 \\ .300243 \\ .400243 \end{bmatrix}$

After 10 fares: $\mathbf{q}^{\mathbf{10}} = T^{\mathbf{10}}\mathbf{q} = \begin{bmatrix} .299998819 \\ .3000005905 \\ .4000005905 \end{bmatrix}$

After 20 fares: $\mathbf{q}^{\mathbf{20}} = T^{\mathbf{20}}\mathbf{q} = \begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix}$

After 30 fares: $\mathbf{q}^{\mathbf{30}} = T^{\mathbf{30}}\mathbf{q} = \begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix}$

It appears as though the state vector is converging towards the vector $\begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix}$. No further multiplication will change the state vector. This is called reaching a *stable distribution* or a *steady state*. Recall that we are iterating $\mathbf{q}^{k+1} = T\mathbf{q}^k$ and eventually for some large k we will reach a point where $\mathbf{q}^{k+1} = \mathbf{q}^k$ and thus $\mathbf{q}^k = T\mathbf{q}^k$ results.

The above example illustrates Markov chains. In a Markov chain, the next state was determined by the transition probabilities and the current state of the system. “A Markov chain is a special … process in which the transition probabilities are constant and independent of the previous behavior of the system.” [20] A Markov chain is an extremely useful tool when we are trying to determine the probability of a system moving (transitioning) from one state to another after *n* steps. According to Carter, Tapia and Papakonstantinou [5, chapter 7],

> “a problem can be considered a (homogeneous) Markov chain if it has the following properties:
>
> a. For each time period, every object (person) in the system is in exactly one of the defined states. At the end of the time period, each

object either moves to a new state or stays in that same state for another period.

b. The objects move from one state to the next according to the transition probabilities, which depend only on the current state (they do not take any previous history into account). The total probability of movement from a state (movement from a state to the same state does count as movement) must equal one.

c. The transition probabilities do not change over time (the probability of going from state *A* to state *B* is the same as it will be at any time in the future)."

The first step in using/creating a Markov chain is to determine the transition matrix, $T = \left[t_{ij} \right]$ such that t_{ij} is the probability of the system moving from state *j* to state *i*. The transition matrix of a Markov chain has the following properties:

1. The matrix must be square. Remember that the each row and each column represent a state. Therefore, the number of rows must equal the number of columns, which must equal the number of states.
2. Since all entries in the matrix represent probabilities, each entry must be between 0 and 1 inclusive.
3. The sum of the entries in any column must be equal to one. The sum of the entries in a column is the sum of the transition probabilities from a state to another state. Since a transition must occur, the sum must be equal to 1.

A matrix has these properties if and only if it is a *stochastic* matrix, i.e. it is a square nonnegative matrix such that each column sum is 1 [13].

As illustrated in NCTM's taxi problem above, the probability of beginning in state *j*, and being in state *i* after *n* steps, is given by the *i, j*th entry of T^n.

To utilize the Markov chain, we also need to know something about the initial conditions. This is usually given as an *initial probability vector*. The initial probability vector of a Markov chain with k states is a k x 1 column matrix, $\mathbf{q} = \begin{bmatrix} q_1 \\ \vdots \\ q_k \end{bmatrix}$, where q_i is the probability that the system is in state i initially [12]. Since **q** is a probability vector, all of its entries must be between 0 and 1 inclusive and $q_1 + q_2 + ... + q_k = 1$.

Given T, the transition matrix, and **q**, the initial probability vector, we can determine the condition of the system after n transitions by the following equation: $\mathbf{q}^{\mathbf{n}} = T^{\mathbf{n}}\mathbf{q}$. Some references refer to $\mathbf{q}^{n}$ as the *state vector after n transitions*.

If, after a certain number of transitions the distribution does not change, we say that the state vector reaches a *stable distribution*. In other words, as the number of transitions, n, gets very large (as $n \to \infty$), $\mathbf{q}^{\mathbf{n}}$ might approach a limiting "*steady state*" vector $\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_k \end{bmatrix}$, where $\mathbf{s} = T\mathbf{s}$ holds true [12]. Entries in the stable state vector are the probabilities of being in each state over the long run no matter what state is the starting state.

In the taxi problem, all initial distributions will eventually lead to the same steady state vector. While different initial distributions will not affect the steady state vector values, they may affect the amount of time required to converge to the stable distribution.

How can we tell if a Markov chain model will converge to a unique steady state vector regardless of initial conditions? Convergence is guaranteed if the transition matrix is a regular matrix. A stochastic matrix is considered to be *regular* if some power of the matrix has only nonzero entries. The taxi problem matrix is obviously regular, but regularity does not require that all entries of the transition matrix itself be nonzero. For example, $B = \begin{bmatrix} .15 & 1 \\ .85 & 0 \end{bmatrix}$ is considered to be a regular matrix since $B^2 = \begin{bmatrix} .8725 & .15 \\ .1275 & .85 \end{bmatrix}$. If all entries in a transition matrix are between 0 and 1 *exclusively,* then all subsequent powers of the transition matrix will be nonzero entries, and convergence is guaranteed [5]. Some authors denote the letter $\boldsymbol{\pi}$ to this steady state vector.

The steady state vector is defined as $\mathbf{s} = \lim_{n \to \infty} \mathbf{q}^n$. Since **s** is independent from initial conditions, it must remain unchanged when transformed by *T*. This makes **s** an *eigenvector*, with an *eigenvalue* equal to 1 [12]. If we have square matrix, *A*, we say λ is an eigenvalue of *A* if there exists a nonzero vector, **x**, such that $A\mathbf{x} = \lambda\mathbf{x}$. We call **x** an eigenvector and (λ, **x**) an eigenpair [12].

In the above taxi problem, we found our steady state vector by using a process called the *power method*, repeated multiplication of the initial vector by the transition matrix. In Section 3 we will show why this always works for a regular matrix. We can also find the steady state vector by finding an eigenvector for eigenvalue 1 and normalizing it to a probability vector.

For the steady state vector of a Markov chain, we wish to solve $\mathbf{s} = T\mathbf{s}$ for **s**. We begin with an elementary method that works for a small transition matrix, such as the one in the taxi problem.

$$T\mathbf{s} = \mathbf{s}$$

$$T\mathbf{s} - \mathbf{s} = 0$$

$$(T - I)\mathbf{s} = 0$$

$$\begin{bmatrix} -.5 & .1 & .3 \\ .2 & -.6 & .3 \\ .3 & .5 & -.6 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Writing out each equation gives:

$$\begin{aligned} -.5s_1 + .1s_2 + .3s_3 &= 0 \\ .2s_1 - .6s_2 + .3s_3 &= 0 \\ .3s_1 + .5s_2 - .6s_3 &= 0 \end{aligned} .$$

To solve this system of equations use the fact that $s_1 + s_2 + s_3 = 1$, since **s** is a state vector. So, $s_1 = 1 - s_2 - s_3$. Substituting for s_1 into the first two equations gives

$$\begin{aligned} -.5(1 - s_2 - s_3) + .1s_2 + .3s_3 &= 0 \\ .2(1 - s_2 - s_3) - .6s_2 + .3s_3 &= 0 \end{aligned}$$

Next distribute and collect like terms:

$$\begin{aligned} -.5 + .6s_2 + .8s_3 &= 0 \\ .2 - .8s_2 + .1s_3 &= 0 \end{aligned}$$

Using linear combinations:

$$\begin{aligned} -.5 + .6s_2 + .8s_3 = 0 \quad &\rightarrow \quad -.5 + .6s_2 + .8s_3 = 0 \\ .2 - .8s_2 + .1s_3 = 0 \quad &\rightarrow \quad -1.6 + 6.4s_2 - .8s_3 = 0 \end{aligned}$$

$$\begin{aligned} -21 + 7s_2 &= 0 \\ 7s_2 &= 21 \\ s_2 &= 0.3 \end{aligned}$$

It follows that $s_3 = 0.4$, and $s_1 = 0.3$. Thus, the steady state vector for the taxi problem is $\mathbf{s} = \begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix}$, as expected.

For a larger transition matrix, a more general method is required. The system $(T - I)\mathbf{s} = 0$ can be solved for **s** by standard matrix techniques that can be implemented on computer software such as *Mathematica* or *Matlab*. The matrix *T* – *I* is reduced to reduced row echelon form (RREF). For the taxi problem, $\text{RREF}(T - I) = \begin{bmatrix} 1 & 0 & -.75 \\ 0 & 1 & -.75 \\ 0 & 0 & 0 \end{bmatrix}$. It is then easy to solve for the unique probability vector by substitution, resulting in $\mathbf{s} = \begin{bmatrix} .3 \\ .3 \\ .4 \end{bmatrix}$.

Not all Markov chains have a unique steady state vector to which the n^{th} state vector converges independent of initial conditions. For example, the transition matrix $T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has eigenvalues of 1 and –1 and has a unique eigenvector for 1 that is a probability vector $\left(\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}\right)$,

but $T^n\mathbf{q}$ does not converge for most probability vectors **q.** In this situation, the probability of ending in a certain state does depend upon the initial conditions or the starting state [20].

One important special type of a non-regular Markov chain is an *absorbing Markov chain*. "A state S_k of a Markov chain is called an *absorbing state* if, once the Markov chain enters the state, it remains there forever. In other words, the probability of leaving the state is zero. This means $p_{kk} = 1$, and $p_{jk} = 0$ for $j \neq k$. A Markov chain is called an absorbing chain if

(i) it has at least one absorbing state; and

(ii) for every state in the chain, the probability of reaching an absorbing state in a finite number of steps is nonzero.

An essential observation for an absorbing Markov chain is that it will eventually enter an absorbing state." [22]

In order to avoid being an absorbing state Markov chain, it should be possible to move from every state to another state. Absorbing Markov Chains are of mathematical interest, but are not relevant to the use of Markov chains by Google.

Section 2

How Markov Chains are used by Google

There are many applications of Markov chains. Some include physics, biology, economics, engineering, and other fields. But perhaps one of the most interesting examples of the use of Markov chains, is the web search engine, Google [23]. Google utilizes a program called PageRank to prioritize the pages found in a search; this is important because a search usually returns far more pages than the searcher is willing to look at. PageRank was developed in 1998 by Larry Page and Sergey Brin when they were Computer Science graduate students at Stanford University. At the time, it was estimated that there were over 150 million web pages [2]. In the spring of 2005, it was estimated that Google made searches among 8.1 billion web pages [14].

How does Google work? First, robot web crawlers are used to find web pages. These pages are indexed and cataloged. Then these pages are assigned PageRank values. These PageRank values are assigned before any user queries are performed. How, then, are the Page Rank values assigned?

Imagine the World Wide Web as a *directed graph*, that is, a finite set of nodes and a set of ordered pairs of nodes representing directed edges between nodes. Each web page is a node and each hyperlink is a directed edge [8]. See figure 2 below, which illustrates a 6-node sample web. Arrows pointing out of a node are outlinks, and arrows pointing towards a node are inlinks.

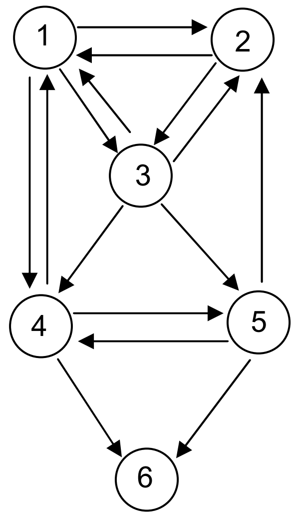


Figure 2: directed graph of sample 6-node web

The numbers of links to (inlinks) and from (outlinks) a page give information about the importance of a page. The more inlinks a web page has, the more important the page. Inlinks from "good" pages carry more weight than inlinks from "weaker", lesser important pages. Also, if a "good" page points to several pages, its weight is distributed proportionally. For example, if an important site like YAHOO! points to your web site but also points to 99 others, you only get credit for $\frac{1}{100}$th of YAHOO!'s PageRank [18].

PageRank begins by defining the rank of a page i by $r_i = \sum_{j \in I_i} \frac{r_j}{|O_j|}$, where r_j is the rank of page j, I_i is the set of pages that point into page i (the number of inlinks to i), and $|O_j|$ is the number of pages that have outlinks from page j. [18]

Notice that this is a recursive definition. To solve, PageRank assigns an initial ranking of $r_i^{(0)} = \frac{1}{n}$, where n is the total number of pages on the web. Then PageRank iterates

$r_i^{(k+1)} = \sum_{j \in I_i} \frac{r_j^{(k)}}{|O_j|}$ for k = 0, 1, 2, …, where $r_i^{(k)}$ is the PageRank of page i at iteration k.

We can write this process using matrix notation. Let $\mathbf{q}^{\mathbf{k}}$ be the PageRank vector at the k^{th} iteration, and let T be the transition matrix for the web; then $\mathbf{q}^{k+1} = T\mathbf{q}^{k}$ [18]. If there are n pages on the web, T is the n x n matrix such that t_{ij} is the probability of moving from page j to page i in one time step. If page j has a set of outlinks, O_j, and we assume all outlinks are equally likely to be chosen,

$$t_{ij} = \begin{cases} \dfrac{1}{|O_j|}, \text{ if there is a link from } j \text{ to } i \\ 0, \text{ otherwise} \end{cases}$$

Using our 6-node sample web from figure 2, we can build the following transition matrix:

$$T = \begin{bmatrix} 0 & 1/2 & 1/4 & 1/3 & 0 & 0 \\ 1/3 & 0 & 1/4 & 0 & 1/3 & 0 \\ 1/3 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/4 & 0 & 1/3 & 0 \\ 0 & 0 & 1/4 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 \end{bmatrix}$$

In the matrix, notice that row i has non-zero elements in positions that correspond to inlinks to page i. Column j has non-zero elements in positions where there are outlinks to page j. If page j has outlinks, the sum of the column is equal to 1 [14].

When we have a column with a sum of zero, that indicates that we have a page with no outlinks. This type of page (node) is sometimes called a dangling node. Dangling nodes represent a problem when trying to set up a Markov model. To alleviate this problem, we replace such columns with $\frac{\mathbf{e}}{n}$, where **e** is a column vector of all ones and n is the order of T. In our 6-node sample web, this creates a new matrix: $\overline{T} = \begin{bmatrix} 0 & 1/2 & 1/4 & 1/3 & 0 & 1/6 \\ 1/3 & 0 & 1/4 & 0 & 1/3 & 1/6 \\ 1/3 & 1/2 & 0 & 0 & 0 & 1/6 \\ 1/3 & 0 & 1/4 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1/4 & 1/3 & 0 & 1/6 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/6 \end{bmatrix}$. The result is a stochastic matrix.

But being stochastic is not enough to guarantee that our Markov model will converge and that a stationary or steady state distribution exists. The other problem facing our transition matrix, $\overline{T}$, and any transition matrix created for the web, is that the matrix may not be regular. The web's nature is such that $\overline{T}$ would not be regular, so we need to make further adjustments. Brin and Page force the transition matrix to be regular by making sure every entry satisfies $0 < t_{ij} < 1$. This guarantees convergence of $\mathbf{q}^n$ to an unique, positive steady state vector, as we will show in Section 3. According to Langville and Meyer, Brin and Page added a perturbation matrix $E = \frac{\mathbf{ee}^T}{n}$ to form what is generally called the "Google Matrix": [17]

$$\overline{\overline{T}} = \alpha\overline{T} + (1-\alpha)E \text{ , for some } 0 \le \alpha \le 1.$$

Google reasoned that this new matrix, $\overline{\overline{T}}$, tends to better model the "real-life" surfer. Such a surfer has a $1-\alpha$ probability of jumping to a random place on the web (i.e. by typing an URL on the command line) and an α probability of deciding to click on an outlink on a current page [14]. The exact value of α remains Google's secret. Many sources report $\alpha \approx 0.85$.

Using 0.85 for our 6-node sample web, we can calculate our "Google matrix", $\overline{\overline{T}}$:

$$\overline{\overline{T}} = 0.85\overline{T} + (1-.85)E$$

$$= 0.85\begin{bmatrix} 0 & 1/2 & 1/4 & 1/3 & 0 & 1/6 \\ 1/3 & 0 & 1/4 & 0 & 1/3 & 1/6 \\ 1/3 & 1/2 & 0 & 0 & 0 & 1/6 \\ 1/3 & 0 & 1/4 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1/4 & 1/3 & 0 & 1/6 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/6 \end{bmatrix} + 0.15\begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

$$= \begin{bmatrix} 1/40 & 9/20 & 19/80 & 37/120 & 1/40 & 1/6 \\ 37/120 & 1/40 & 19/80 & 1/40 & 37/120 & 1/6 \\ 37/120 & 9/20 & 1/40 & 1/40 & 1/40 & 1/6 \\ 37/120 & 1/40 & 19/80 & 1/40 & 37/120 & 1/6 \\ 1/40 & 1/40 & 19/80 & 37/120 & 1/40 & 1/6 \\ 1/40 & 1/40 & 1/40 & 37/120 & 37/120 & 1/6 \end{bmatrix}$$

Calculating powers of the transition matrix is one way that we can determine the stationary vector. Since $\left(\overline{\overline{T}}\right)^{25} = \begin{bmatrix} .2066 & .2066 & .2066 & .2066 & .2066 & .2066 \\ .1770 & .1770 & .1770 & .1770 & .1770 & .1770 \\ .1773 & .1773 & .1773 & .1773 & .1773 & .1773 \\ .1770 & .1770 & .1770 & .1770 & .1770 & .1770 \\ .1314 & .1314 & .1314 & .1314 & .1314 & .1314 \\ .1309 & .1309 & .1309 & .1309 & .1309 & .1309 \end{bmatrix}$ (to four significant digits), we can deduce that $\left(\overline{\overline{T}}\right)^{k}\mathbf{q}$ is converging to the values shown in each column.

The stationary vector for our 6-node sample web would be $\mathbf{s} = \begin{bmatrix} .2066 \\ .1770 \\ .1773 \\ .1770 \\ .1314 \\ .1309 \end{bmatrix}$.

How does Google or PageRank use this stationary vector? Suppose a user enters a query in the Google search window. Suppose the query requests term 1 or term 2. “The inverted term-document file below is accessed. (Inverted file storage is similar to the index in the back of a book – it is a table containing a row for every term in a collection's dictionary. Next to each term is a list of all documents that use that term)." [17]

Assume the inverted file storage of document content is:

term 1 $\rightarrow$ doc 2, doc 5, doc 6
term 2 $\rightarrow$ doc 2, doc 3
$\vdots$

Thus, the relevancy set for the user's query is {2, 3, 5, 6}. The PageRank of these four documents is now compared to determine order of importance. According to our 6-node model, $s_2 = .1770, s_3 = .1773, s_5 = .1314,$ and $s_6 = .1309$. Thus, PageRank deems document 3 most important, followed by document 2, document 5, and document 6. When a new query is entered, the inverted term document file is accessed again and a new relevancy set is created.

According to the Google website, PageRank continues to be "the heart of our software" [10]. Keep in mind, however, that PageRank is not the only criteria Google uses to categorize the importance of a web page. Since Google is a full text search engine, it utilizes PageRank as well as a number of other factors to rank search results. These other factors include "standard information retrieval measures, proximity, and anchor text" [2].

PageRank is an interesting application of Markov chains; one that would be of interest to the "web-surfing" high school student population. It would be a great extension of the matrix chapter to incorporate a small unit on Markov chains. Included in the Appendix are two sample worksheets that could be used for a unit on Markov chains.

Section 3

The Mathematics of Markov Chains

According to Hogben, a "*Markov chain* is a random process described by a physical system that at any given time t = 1, 2, 3, … occupies one of a finite number of states. At each time t the system moves from state j to state i with probability p_{ij} that does not depend on t. The numbers p_{ij} are called *transition probabilities*." [12] A notable feature of Markov chains is that they are historyless – the next state of the system depends only on the current state, not any prior states [11].

A key component of a Markov chain is its transition matrix. A *transition matrix*, T, of a Markov chain is an n x n matrix, where n is the number of states of the system. Each entry of the transition matrix, t_{ij}, is equal to the probability of moving from state j to state i in one time interval. Thus, $0 \le t_{ij} \le 1$ must be true for all i, j = 1, 2, … n. An example of a transition matrix was found in Section 1 with our taxi problem: $T = \begin{bmatrix} .5 & .1 & .3 \\ .2 & .4 & .3 \\ .3 & .5 & .4 \end{bmatrix}$. From this transition matrix, we can quickly see the probability of moving from one state to another. For example, we note that $t_{3,2}$ = 0.5; this tells us there is a 50% probability of moving from state 2 to state 3. Since a system must be in one state or another, the sum of each column in a transition matrix must be equal to 1.

Besides the transition matrix, we need to know what state the system is currently in. This information is given in an n x 1 matrix and is called a *state vector*, $\mathbf{q} = \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix}$. The state a system is in initially is called an *initial state vector*. If the system is currently in state j, q_j = 1, while all other entries of the initial state vector will be equal to zero. If the initial state information is not known, but the likelihood of the system being in a certain state is, we use an *initial probability vector* instead of an initial state vector. If **q** is to be a probability vector, two conditions must hold:

1. $0 \le q_i \le 1$, since each entry is a probability.
2. $q_1 + q_2 + ... + q_n = 1$, because the system must be in one of the *n* states at any given time.

Given an initial probability vector, the k^{th}*-step probability vector* is the vector $\mathbf{q}^k = \begin{bmatrix} q_1^{\ k} \\ \vdots \\ q_n^{\ k} \end{bmatrix}$, where q_i^k is the probability of being in state *i* after *k* steps. We can find the k^{th} step probability vector, $\mathbf{q}^k$, by using the transition matrix, *T* and the following theorem.

Theorem 1

(a) Let *T* be a transition matrix of a Markov chain. Then the probability of moving from state *j* to state *i* in *k* time increments (*k* > 0 and *k* must be an integer) is given by the i,j^{th} entry of the matrix T^k.

(b) If *T* is a transition matrix of a Markov chain with an initial probability vector **q**, and a k^{th}-step probability vector $\mathbf{q}^k$, then $\mathbf{q}^k = T^k\mathbf{q}$. [12, pp. 84-86]

Proof of part (b), by induction: Start with k = 1. The 1^{st} step probability vector, $\mathbf{q}^1$, would be found by multiplying the transition matrix by the initial probability vector, $\mathbf{q}^1 = T\mathbf{q}$. Now assume that for some positive integer *k*, $\mathbf{q}^k = T^k\mathbf{q}$ holds true. We wish to show that $\mathbf{q}^{k+1} = T^{k+1}\mathbf{q}$ is true. To get to the $(k+1)^{th}$ step, we would first need to get the k^{th} step by $\mathbf{q}^k = T^k\mathbf{q}$ and then proceed one more step.

$$\begin{aligned} \mathbf{q}^{k+1} &= T\mathbf{q}^{k} \\ &= TT^{k}\mathbf{q} \quad \text{by our assumption} \\ &= T^{k+1}\mathbf{q} \end{aligned}$$

Thus for any positive integer, $\mathbf{q}^k = T^k\mathbf{q}$ is true.

Part (a) follows from (b) by using the i^{th} standard vector $\mathbf{e}_i$ as the initial vector, which means the system is in state *i* initially.

Often when dealing with Markov chains, we are interested in what is happening to the system in the long run or as $k \to \infty$. As *k* approaches infinity, $\mathbf{q}^k$ will often approach a limiting vector, $\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$, called a *steady state vector*. A matrix *A* is called *regular* when for some positive integer *k*, all entries of A^k are positive. We will prove that if a transition matrix *T* is regular, then there is a unique steady state vector **s** and $\mathbf{q}^k \to \mathbf{s}$ as $k \to \infty$, for any initial probability vector, **q**.

Back in Section 1, we observed that our steady state vector is an eigenvector for eigenvalue 1. If *A* is an *n* x *n* square matrix, a number λ is called an *eigenvalue* of *A* if there exists a non-zero column vector **x** for which $A\mathbf{x} = \lambda\mathbf{x}$. The column vector, **x**, is called an *eigenvector* of *A* (for the eigenvalue λ). Eigenvalues are also called *characteristic values* or *characteristic roots* [12]. When dealing with Markov chains, we are interested in the long-term behavior or steady state vector; i.e. we are interested in what happens when $\lim_{k\to\infty} T^k\mathbf{q} = \mathbf{s}$. The eigenvalue 1 plays a vital role.

Theorem 2 If T is a transition matrix and the $\lim_{k \to \infty} T^k \mathbf{q} = \mathbf{s}$ then $T\mathbf{s} = \mathbf{s}$ and $\mathbf{s}$ is an eigenvector for eigenvalue 1.

Proof: If the $\lim_{k \to \infty} T^k \mathbf{q} = \mathbf{s}$, we can multiply both sides on the left by T. Then we have $\lim_{k \to \infty} TT^k \mathbf{q} = T\mathbf{s}$ or $\lim_{k \to \infty} T^{k+1} \mathbf{q} = T\mathbf{s}$. The limit as k approaches infinity for T^{k+1} is no different than the limit as k approaches infinity for T^k. Thus,

$\lim_{k \to \infty} T^{k+1} \mathbf{q} = \lim_{k \to \infty} T^k \mathbf{q} = \mathbf{s}$ and $\mathbf{s} = T\mathbf{s}$.

Given a value λ, and a vector **v**, it is easy to determine by computation whether **v** and λ form an eigenvector eigenvalue pair of matrix A. Often, however, we are not given the eigenvalues and eigenvectors of a system, but must instead find them. We begin with the equation $A\mathbf{x} = \lambda\mathbf{x}$. Use the property that $I\mathbf{x} = \mathbf{x}$, where I is the identity matrix. Rewrite $A\mathbf{x} = \lambda\mathbf{x}$ in the form of $A\mathbf{x} = \lambda I\mathbf{x}$ or $A\mathbf{x} - \lambda I\mathbf{x} = 0$. A more useful form would then be $(A - \lambda I)\mathbf{x} = 0$. Since we required **x** to be a non-zero column vector, $(A - \lambda I)\mathbf{x} = 0$ will have a non-zero solution if and only if $\det(A - \lambda I) = 0$. Thus λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$ [12]. The equation $\det(A - \lambda I) = 0$ is called the *characteristic equation*. If we can solve the characteristic equation, we can find the eigenvalues of matrix A. If A is an n x n matrix, there will be n eigenvalues of matrix A over the complex numbers. Some of these eigenvalues might be repeated (multiple roots of the characteristic equation) [5, chapter 9]. A *simple eigenvalue* is an eigenvalue that has the property that it is not a repeated root.

Lemma 3 For any square matrix A, λ is an eigenvalue of A if and only if λ is an eigenvalue of A^T.

Proof: The remarks in the previous paragraph establish λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$. For any matrix B, $\det B = \det B^T$, so $0 = \det(A - \lambda I)$ if and only if $0 = \det((A - \lambda I)^T) = \det(A^T - \lambda I)$. Thus λ is an eigenvalue of A if and only if λ is an eigenvalue of A^T.

Theorem 4 If every column sum of an n x n matrix A is 1, then 1 is an eigenvalue of matrix A.

Proof: Consider the n x n matrix A, where each column sum is 1. By Lemma 3, it is sufficient to show that 1 is an eigenvalue of A^T. Since each column sum of A, is equal to 1, each row sum of A^T is 1 and $A^T\mathbf{e} = \mathbf{e}$, where **e** is a column vector of all ones. Therefore, A^T has an eigenvalue of 1 and so A has an eigenvalue of 1.

Before we continue, we need to define a few more terms. First, a *nonnegative matrix* is a matrix whose entries are nonnegative. That is, for all entries, a_{ij} λ 0. Second, a *positive* matrix is a matrix whose entries are positive. In this case all entries, $a_{ij} > 0$. Third, recall that a *stochastic matrix* is a square nonnegative matrix, with the entries in each column summing to 1. A Markov transition matrix is an example of a stochastic matrix. Last of all, the *spectral radius,* ρ*,* of a matrix is the largest absolute value of an eigenvalue of a matrix.

Theorem 5 If A is stochastic, then every eigenvalue λ of A satisfies $|\lambda| \le 1$.

Proof: Consider A, an n x n matrix with columns that each sum to 1. Let $B = A^T$, so B has row sums of 1. By Lemma 3, it is enough to show that $|\lambda| \le 1$ for *B.* . Let

$B\mathbf{v} = \lambda\mathbf{v}$, with nonzero vector $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ (λ is an eigenvalue). Let *m* be such that

$|v_i| \le |v_m|$ for all *i*.

$$\begin{aligned} |\lambda||v_m| = |\lambda v_m| = |(B\mathbf{v})_m| = |b_{m1}v_1 + b_{m2}v_2 + \cdots + b_{mn}v_n| \le |b_{m1}||v_1| + |b_{m2}||v_2| + \cdots |b_{mn}||v_n| \\ \le |b_{m1}||v_m| + |b_{m2}||v_m| + \cdots |b_{mn}||v_m| \le |v_m| \end{aligned}$$

Since **v** is nonzero, we can divide both sides by $|v_m|$, giving $|\lambda| \le 1$. Thus *B*, and ultimately *A*, have eigenvalues such that every eigenvalue λ satisfies $|\lambda| \le 1$.

Every Markov transition matrix has columns that sum to 1. Using the above theorems allows us to conclude the following about a Markov transition matrix:

1. The transition matrix will have an eigenvalue of 1.
2. Other eigenvalues λ of the transition matrix must satisfy $|\lambda| \le 1$.

The next corollary follows from Theorems 4 and 5.

Corollary 6 If *T* is a transition matrix then $\rho(T) = 1$.

The next two theorems are stated below without proof. The second theorem, Theorem 8, is clear if matrix *A* is diagonalizable. However, we wish to use it in a more general case and therefore state it without proof.

Theorem 7 If *A* is a positive matrix then

i) $\rho(A)$ is a simple eigenvalue > 0

ii) if λ is an eigenvalue not equal to $\rho(A)$, then $|\lambda| < \rho(A)$

iii) $\rho(A)$ has a positive eigenvector **x**, and any eigenvector for $\rho(A)$ is a multiple of **x**. [13]

Theorem 8 Let A be an n x n real or complex matrix having eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ (not necessarily distinct). Then for any positive integer k, eigenvalues of A^k are $\lambda_1{}^k, \lambda_2{}^k, ..., \lambda_n{}^k$. [13]

Using Theorems 7 and 8 allows us to prove the following theorem:

Theorem 9 If T is a regular transition matrix, then there is a unique probability vector **s** that is an eigenvector for eigenvalue 1, and if λ is an eigenvalue not equal to 1 then $|\lambda| < 1$.

Proof: Let T be a regular transition matrix. Therefore, from Corollary 6, we know $\rho(T) = 1$. Since T is regular, some T^k will be positive by the definition of a regular matrix, and then from Theorem 7 we know the following:

1) $\rho(T^k) = 1$ is a simple eigenvalue of T^k,
2) If μ is any eigenvalue of T^k and $\mu \neq 1$, then $|\mu| < 1$,
3) Eigenvalue 1 of T^k has a positive eigenvector, **x**, and any eigevectors for 1 is a multiple of **x.**

By Theorem 8, all the eigenvalues of T^k are eigenvalues of T raised to the k^{th} power, so

1) 1 is a simple eigenvalue of T and
2) if $\lambda \neq 1$ is an eigenvalue of T then $|\lambda| < 1$.

If **x** is an eigenvector for 1, $T\mathbf{x} = 1\mathbf{x} = \mathbf{x}$. Since 1 is a simple eigenvalue of T^k, up to multiplication by a scalar there is a unique eigenvector for 1, so any eigenvector of T^k is an eigenvector of T and T has a positive eigenvector for 1. If we take any positive eigenvector for 1 and divide it by the sum of its entries, we obtain the unique eigenvector for 1 that is a probability vector; we will call this vector **s**.

Back in Theorem 1, we proved that $\mathbf{q}^k = T^k\mathbf{q}$. The computation of $T^k\mathbf{q}$ is the power method, and we can rewrite this equation as $\mathbf{q}^k = T^k\mathbf{q} = c_1\lambda_1^k\mathbf{x_1} + c_2\lambda_2^k\mathbf{x_2} + ... + c_n\lambda_n^k\mathbf{x_n}$, where $\mathbf{q} = c_1 g x_1 + ... + c_n g x_n$ [21]. Note that we are assuming that there are n independent eigenvectors and that $c_1 \neq 0$. By Corollary 6 and Theorem 7(ii), $\rho(T) = 1 = \lambda_1$ and that all other eigenvalues, $|\lambda| < 1$. As $k \to \infty$, $(c_2\lambda_2^k\mathbf{x_2} + ... + c_n\lambda_n^k\mathbf{x_n}) \to 0$, since $|\lambda| < 1$, leaving us with $\lim_{k\to\infty}\mathbf{q}^k = T^k\mathbf{q} = c_1\mathbf{x_1}$. So the steady state vector, **s**, is the eigenvector $c_1\mathbf{x_1}$, for eigenvalue 1. We can then rewrite our equation as $\lim_{k\to\infty}\mathbf{q}^k = \lim_{k\to\infty}T^k\mathbf{q} = \mathbf{s}$. We will show that the assumptions that there are n independent eigenvectors and that $c_1 \neq 0$ are unnecessary.

The following theorem is clearer if you have a diagonalizable matrix. Since we want the more general case, we will state it without proof.

Theorem 10 If A is a positive matrix and $\rho(A) < 1$, then $\lim_{k\to\infty}A^k = 0$. [13]

Back in section 1, we observed that taking larger and larger powers of our transition matrix, T, seemed to show T^k converging towards a matrix whose columns were the steady state vector, **s**. Theorem 11 below proves that if the transition matrix is regular, it will converge to $[\mathbf{s} \quad \mathbf{s} \quad ... \quad \mathbf{s}]$.

Theorem 11 If T is a regular stochastic matrix then $\lim_{k\to\infty} T^k = [\mathbf{s} \quad \mathbf{s} \quad \mathbf{...} \quad \mathbf{s}]$, where **s** is the unique probability vector that is an eigenvector for eigenvalue 1.

Proof: Let T be a regular stochastic matrix. By Theorem 9, T has a unique probability eigenvector, **s**, for 1 so $T\mathbf{s} = \mathbf{s}$. Let $L = [\mathbf{s} \quad \mathbf{s} \quad \mathbf{...} \quad \mathbf{s}]$.

$[1 \quad ... \quad 1]T = [1 \quad ... \quad 1]$, since T has a column sum of 1.

$TL = [T\mathbf{s} \quad T\mathbf{s} \quad ... \quad T\mathbf{s}] = [\mathbf{s} \quad \mathbf{s} \quad ... \quad \mathbf{s}] = L$

$L = \mathbf{s}[1 \quad 1 \quad ... \quad 1]$

$LT = \mathbf{s}[1 \quad 1 \quad ... \quad 1]T = \mathbf{s}[1 \quad 1 \quad ... \quad 1] = L$

$L^2 = \mathbf{s}[1 \quad 1 \quad ... \quad 1]\mathbf{s}[1 \quad 1 \quad ... \quad 1] = \mathbf{s}[1 \quad 1 \quad ... \quad 1] = L$, since $[1 \quad 1 \quad ... \quad 1]\mathbf{s} = 1$.

$(T-L)L = TL - L^2 = L - L = 0$. Likewise, $L(T-L) = 0$.

$(T-L)^k = T^k - L$

Proof of $(T-L)^k = T^k - L$ (by induction): Let k = 1.

$(T-L)^1 = T - L = T^1 - L$. Assume for some k, $(T-L)^k = T^k - L$ is true.

We wish to show that for k + 1, $(T-L)^{k+1} = T^{k+1} - L$ is true.

$(T-L)^{k+1} = (T-L)^k(T-L) = (T^k - L)(T-L)$ by our assumption.

$$= T^k(T-L) - L(T-L) = T^k(T-L) - 0 = T^kT - T^kL$$
$$= T^{k+1} - L$$

Thus, $(T-L)^k = T^k - L$ holds true for all positive integers, k.

Now we show every nonzero eigenvalue of $(T - L)$ is an eigenvalue of L.

Suppose $(T-L)\mathbf{w} = \mu\mathbf{w}$, $\mu \neq 0$ is eigenvalue, $\mathbf{w} \neq 0$ is eigenvector.

$\mu L\mathbf{w} = L(\mu\mathbf{w}) = L(T - L)\mathbf{w} = 0\mathbf{w} = 0$. Because $\mu \neq 0$, $L\mathbf{w} = 0$. Thus,

$\mu\mathbf{w} = (T - L)\mathbf{w} = T\mathbf{w} - L\mathbf{w} = T\mathbf{w}$ and λ is eigenvalue for T with eigenvector $\mathbf{w}$.

1 is not an eigenvalue of $(T - L)$ because $(T - L)\mathbf{s} = T\mathbf{s} - L\mathbf{s} = \mathbf{s} - \mathbf{s} = 0$. Thus any nonzero eigenvalue λ of $(T - L)$ satisfies $|\mu| < 1$. So $\rho(T - L) < 1$ and by Theorem 10, $\lim_{k\to\infty}(T - L)^k = 0$. Since $(T - L)^k = T^k - L$, $\lim_{k\to\infty}(T^k - L) = 0$.

Therefore, as k approaches infinity, $\lim_{k\to\infty} T^k = L = [\mathbf{s} \quad \mathbf{s} \quad \mathbf{...} \quad \mathbf{s}]$.

We have now established and can conclude with the following theorem:

Theorem 12 If a Markov chain has a regular Markov transition matrix T, then there is a unique probability vector **s** such that $T\mathbf{s} = \mathbf{s}$. Furthermore, **s** is the steady state vector for any initial probability vector, i.e. for any initial probability vector **q** the sequence of vectors **q**, $T\mathbf{q}$, $T^2\mathbf{q}$, …$T^k\mathbf{q}$, … approaches **s**.

[1] Blake, Lewis. (1998). “Markov chains” Retrieved 4/13/05 from http://www.math.duke.edu/education/ccp/materials/linalg/markov/index.html

[2] Brin, Sergey; Page, Larry; Motwami, R.; Winograd, Terry. “The PageRank Citation Ranking: Bringing Order to the Web”, January 29, 1998. http://citeseer.ist.psu.edu/page98pagerank.html

[3] Bryant, David. (2004). “Mathematics and Statistics 240 - Chapter 5.1 and 5.5 Notes”. Retrieved 6/3/05 from http://www.mcb.mcgill.ca/~bryant/MS240

[4] Burkhadt, Jim. (2003). “Linear Algebra Glossary”. Retrieved 6/4/05 and 6/5/05 from http://www.csit.fsu.edu/~burkhardt/papers/linear_glossary.html

[5] Carter, Tamara; Tapia, Richard A.; Papakonstantinou, Anne. (1995-2003). “Linear Algebra: An Introduction to Linear Algebra for Pre-Calculus Students”, chapter 7 and 9, Retrieved 4/20/05 from http://ceee.rice.edu/Books/LA/index.html

[6] Chien, Dwork, Kumar, and Sivakumar. (2001).“Towards exploiting link evolution”. Retrieved 6/2/05 from CiteSeer, http://citeseer.ist.psu.edu/chien01towards.html

[7] Demmel, Jim. (1996). “CS 267 – Applications of Parallel Computers – Notes from Lecture 23(b)”. Retrieved 6/3/05 from http://www.cs.berkely.edu/~demmel/cs267/lecture20/lecture20b.html

[8] Donnelly, Karen E. (no date). “M244 Linear Algebra: Markov Chains, Winter 2005”. Retrieved 5/23/05 from http://www.saintjoe.edu/~karend/m244/NBK6-1.html

[9] Durrett, R. Math 275 Course, Chapter 5, Retrieved 4/20/05 from http://www.math.cornell.edu/%7Edurrett/math275/ch5.pdf

[10] Google Technology. (2004). Retrieved 4/20/05 from http://www.google.com/technology/index.html

[11] Hefferon, Jim. (2003). “Linear Algebra”, chapter 3. Retrieved 4/20/05 from http://joshua.smcvt.edu/linearalgebra

[12] Hogben, Leslie. *Elementary Linear Algebra*. West Publishing Company, St. Paul, MN, 1987: 81-92

[13] Horn, R.; Johnson, C. R. *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.

[14] Hyvönen, Saara. (2005). "Linear Algebra Methods for Data Mining". Retrieved 5/15/05 from http://www.cs.helsinki.fi/u/szhyvone/lecture9.pdf

[15] Kemeny, John; Snell, J. Laurie; Thompson, Gerald L. "Introduction to Finite Mathematics", 3rd Edition. Retrieved 4/6/05 from http://math.dartmouth.edu/~doyle/docs/finite/cover/cover.html

[16] Langville, Amy N.; Meyer, Carl D. "Fiddling with PageRank", August 15, 2003. Retrieved 6/2/05 from CiteSeer, http://citeseer.ist.psu.edu/713721.html

[17] Langville, Amy N.; Meyer, Carl D. "A Survey of Eigenvector Methods of Web Information Retrieval", January 5, 2004. Retrieved 6/2/05 from CiteSeer, http://citeseer.ist.psu.edu/636124.html

[18] Langville, Amy N.; Meyer, Carl D. "The Use of the Linear Algebra by Web Search Engines", October 19, 2004. Retrieved 6/2/05 from CiteSeer, http://citeseer.ist.psu.edu/718721.html

[19] Lipschutz, Seymour. *Linear Algebra*. Second Edition. McGraw-Hill, Inc., New York, 1991

[20] National Council of Teachers of Mathematics (NCTM). *New Topics for Secondary School Mathematics: Matrices*. Reson, VA: NCTM, 1988: 1, 2, 49-65, 96-98

[21] Strang, Gilbert. Professor Strang's Class 18.06 Linear Algebra Lecture Videos – Lectures #21 and #24, Fall 1999. Retrieved 4/15/05 from http://ocw.mit.edu/OcwWeb/Mathematics/18-06Linear-AlgebraFall2002/VideoLectures/index.htm

[22] Weckesser, Warren. (2005). "Math 312 Lecture Notes: Markov Chains". Retrieved 6/25/05 from http://math.colgate.edu/~wweckesser/math312/handouts/MarkovChains.pdf

[23] Wikipedia. (2005). “Examples of Markov chains”. Retrieved 4/21/05 from Wikipedia, http://en.wikipedia.org/wiki/Examples_of_Markov_chains

Appendix

Markov Chain Worksheet #1

Name ____________________ P. ____

1. In the Quad Cities, customers can choose from three major grocery stores: H-Mart, Freddy's and Shopper's Market. Each year H-Mart retains 80% of its customers, while losing 5% to Freddy's and 15% to Shopper's Market. Freddy's retains 65% of its customers, loses 20% to H-Mart and 15% to Shopper's Market. Shopper's Market keeps 70% of its customers, loses 20% to H-Mart and 10% to Freddy's. Currently, let us assume that H-mart has ½ of the market, and Shopper's Market and Freddy's each have ¼ of the market.

   a) Draw a state diagram representing this problem.

   b) Set up a transition matrix and write down the initial state vector.

H-Mart | Shopper's Market | Freddy's

$$T = \begin{matrix} \text{H-Mart} \\ \text{Shopper's Market} \\ \text{Freddy's} \end{matrix} \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \qquad \mathbf{q} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

   c) Use the transition matrix and initial state vector to complete the table showing the market share for the next ten years. (Use 4 decimal places where necessary.)

	1	2	3	4	5	6	7	8	9	10
H-Mart										
Shopper's Market										
Freddy's										

d) Compute $T^{n}\mathbf{q}$ for a large enough n to determine the steady state vector to 4 decimal places. Write down T^{n} and the n you used.

$$T^{n} = \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$$ n = ____________

e) What did you notice about the columns of T^{n} and the steady state vector?

f) Make up two other initial state vectors (Remember: all entries must be ≥ 0 and their sum must be equal to 1). Determine the steady state vector in the same manner as you did above. Record your information and findings below.

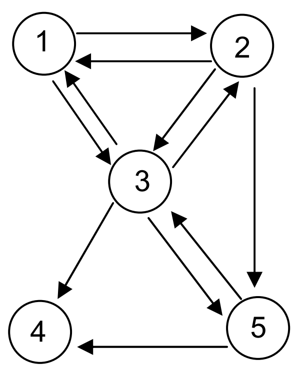
i) $\mathbf{q} = \begin{bmatrix} \quad \\ \quad \end{bmatrix}$ steady state vector $= \begin{bmatrix} \quad \\ \quad \end{bmatrix}$ Power of T used: ________

ii) $\mathbf{q} = \begin{bmatrix} \quad \\ \quad \end{bmatrix}$ steady state vector $= \begin{bmatrix} \quad \\ \quad \end{bmatrix}$ Power of T used: ________

Markov Chain Worksheet #2

Name ____________________ P. ____

The following is a sample 5-page web:



Create a transition matrix for our web:

$T =$

Modify our transition matrix, if necessary, to eliminate any dangling nodes:

$\overline{T} =$

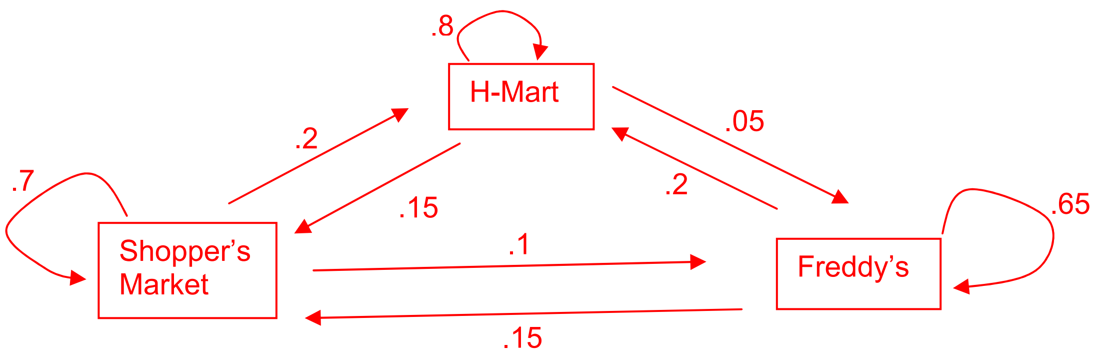
Further modify $\overline{T}$ to create our "Google Matrix". Use $\overline{\overline{T}} = \alpha\overline{T} + (1-\alpha)E$, where $E =$ matrix with all entries $= \frac{1}{5}$. Allow $\alpha = 0.9$.

Markov Chain Worksheet #1

Name ___KEY_______________ P. _____

1. In the Quad Cities, customers can choose from three major grocery stores: H-Mart, Freddy's and Shopper's Market. Each year H-Mart retains 80% of its customers, while losing 5% to Freddy's and 15% to Shopper's Market. Freddy's retains 65% of its customers, loses 20% to H-Mart and 15% to Shopper's Market. Shopper's Market keeps 70% of its customers, loses 20% to H-Mart and 10% to Freddy's. Currently, let us assume that H-mart has ½ of the market, and Shopper's Market and Freddy's each have ¼ of the market.

   a) Draw a state diagram representing this problem.



   b) Set up a transition matrix and write down the initial state vector.

$$T = \begin{array}{r} \text{H-Mart} \\ \text{Shopper's Market} \\ \text{Freddy's} \end{array} \overset{\begin{array}{ccc}\text{H-Mart} & \text{Shopper's Market} & \text{Freddy's}\end{array}}{\begin{bmatrix} .80 & .20 & .20 \\ .15 & .70 & .15 \\ .05 & .10 & .65 \end{bmatrix}} \qquad \mathbf{q} = \begin{bmatrix} 1/2 \\ 1/4 \\ 1/4 \end{bmatrix}$$

   c) Use the transition matrix and initial state vector to complete the table showing the market share for the next ten years. (Use 4 decimal places where necessary.)

	1	2	3	4	5	6	7	8	9	10
H-Mart	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
Shopper's Market	.2875	.3081	.3195	.3257	.3291	.3310	.3321	.3326	.3329	.3331
Freddy's	.2125	.1919	.1805	.1743	.1709	.1690	.1679	.1673	.1671	.1669

d) Compute $T^n\mathbf{q}$ for a large enough n to determine the steady state vector to 4 decimal places. Write down T^n and the n you used.

$$T^n = \begin{bmatrix} .5000 & .5000 & .5000 \\ .3333 & .3333 & .3333 \\ .1667 & .1667 & .1667 \end{bmatrix}$$

n = _____20 or more_____

e) What did you notice about the columns of T^n and the steady state vector?

Each column is equal to the steady state vector.

f) Make up two other initial state vectors (Remember: all entries must be ≥ 0 and their sum must be equal to 1). Determine the steady state vector in the same manner as you did above. Record your information and findings below. ANSWERS WILL VARY. Steady state vector should remain the same.

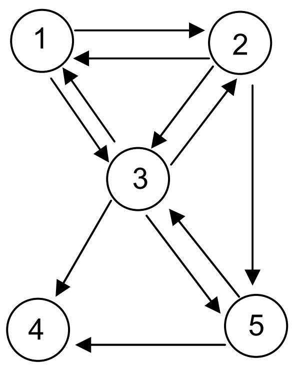
i) $\mathbf{q} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$ steady state vector $= \begin{bmatrix} \\ \\ \\ \end{bmatrix}$ Power of T used: _________

ii) $\mathbf{q} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$ steady state vector $= \begin{bmatrix} \\ \\ \\ \end{bmatrix}$ Power of T used: _________

Markov Chain Worksheet #2

Name _____KEY_______________ P. _____

The following is a sample 5-page web:



Create a transition matrix for our web:

$$T = \begin{bmatrix} 0 & 1/3 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 0 & 0 \\ 1/2 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 1/4 & 0 & 1/2 \\ 0 & 1/3 & 1/4 & 0 & 0 \end{bmatrix}$$

Modify our transition matrix, if necessary, to eliminate any dangling nodes:

$$\overline{T} = \begin{bmatrix} 0 & 1/3 & 1/4 & 1/5 & 0 \\ 1/2 & 0 & 1/4 & 1/5 & 0 \\ 1/2 & 1/3 & 0 & 1/5 & 1/2 \\ 0 & 0 & 1/4 & 1/5 & 1/2 \\ 0 & 1/3 & 1/4 & 1/5 & 0 \end{bmatrix}$$

Further modify $\overline{T}$ to create our "Google Matrix". Use $\overline{\overline{T}} = \alpha\overline{T} + (1-\alpha)E$, where $E =$ matrix with all entries $= \frac{1}{5}$. Allow $\alpha = 0.9$.

$$\overline{T} = 0.9\begin{bmatrix} 0 & 1/3 & 1/4 & 1/5 & 0 \\ 1/2 & 0 & 1/4 & 1/5 & 0 \\ 1/2 & 1/3 & 0 & 1/5 & 1/2 \\ 0 & 0 & 1/4 & 1/5 & 1/2 \\ 0 & 1/3 & 1/4 & 1/5 & 0 \end{bmatrix} + 0.1\begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}$$

$$= \begin{bmatrix} 1/50 & 8/25 & 49/200 & 1/5 & 1/50 \\ 47/100 & 1/50 & 49/200 & 1/5 & 1/50 \\ 47/100 & 8/25 & 1/50 & 1/5 & 47/100 \\ 1/50 & 1/50 & 49/200 & 1/5 & 47/100 \\ 1/50 & 8/25 & 49/200 & 1/5 & 1/50 \end{bmatrix}$$