

COMP4121 Lecture Notes  
**程序代写代做 CS编程辅导**  
The PageRank, Markov chains and the Perron Frobenius theory



Aleks Ignjatovic

@cse.unsw.edu.au

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

THE UNIVERSITY OF  
NEW SOUTH WALES



School of Computer Science and Engineering  
The University of New South Wales  
Sydney 2052, Australia

# Topic One: the PageRank

Please read Chapter 3 of the textbook Networked Life entitled “How does Google rank webpages?”; other references on the material covered are listed at the end of these lecture notes.

## 1 Problem: ordering webpages according to their importance

**Setup:** Consider all the web pages  $\{P_i : P_i \in \text{WWW}\}$  in the **WWW** (“World Wide Web”) as a directed graph whose nodes are the web pages  $\{P_i : P_i \in \text{WWW}\}$ . There is a directed edge  $P_i \rightarrow P_j$  just in case page  $P_i$  points to page  $P_j$ , i.e. page  $P_i$  has a link to page  $P_j$ .

**Problem:** Rank all the web pages according to their “importance”.<sup>1</sup>

Intuitively, one might point to a page  $P_0$ , then  $P_0$  should have a high rank, because if one includes on their webpage a link to page  $P_0$ , then this can be seen as their recommendation of  $P_0$ .

However, this is not a good criterion for several reasons. For example, it can be easily manipulated to increase the rating of any webpage  $P_0$ , simply by creating a lot of silly web pages which just point to  $P_0$ ; also, if a webpage “generously” points to a very large number of webpages, such “easy to get” recommendation is of dubious value. Thus, we need to refine our strategy how to rank webpages according to their “importance”, by doing something which cannot be easily manipulated “locally” (i.e., by any group of people who might collude, even if such a group is sizeable). One of the crucial insights of Google founders was that the ranks of a webpage should depend on the entire global structure of the web; in other words, we should make the rank of every webpage dependent on the ranks of all other webpages on the entire WWW! Such a method would be virtually immune to spamming, because any group of people can alter only a negligible part of the entire WWW structure.

**Notation:** Assume that  $\rho$  is some ranking of webpages (obtained in whatever way); so with such a ranking function the rank of a page  $P$  is  $\rho(P)$ . Let also  $\#(P)$  be the number of all outgoing links on a webpage  $P$ . As we have mentioned, we will write  $P_1 \rightarrow P_0$  to denote that a page  $P_1$  has a link pointing to a page  $P_0$ .

Intuitively, we would want a web page  $P$  to have a high rank only if it is pointed at by many pages  $P_i$  which themselves have a high rank, and which do not point to an excessive number of other web pages. One idea how to obtain ranking with such a desirable property would be to look for a ranking function  $\rho$  which satisfies

$$\rho(P) = \sum_{P_i \rightarrow P} \frac{\rho(P_i)}{\#(P_i)}, \quad (1.1)$$

for all web pages  $P$  on the web. If (1.1) is satisfied, then a page  $P$  will have a high rank only if the right-hand side of this equation is large. Clearly, the sum on the right is large only if there are enough pages  $P_i$  which point to  $P$  and which: (i) themselves have a high rank  $\rho(P_i)$ ; (ii) do not point to too many other pages besides  $P$ , so that  $\#(P_i)$  is not too large; otherwise  $\rho(P_i)/\#(P_i)$  might be small even if  $\rho(P_i)$  is reasonably large.

Note that the above formula cannot be directly used to compute  $\rho(P)$  because we do not know what  $\rho(P_i)$  are; it is only a condition which  $\rho$  should satisfy. But:

1. why should such  $\rho$  exist at all?
2. even if such  $\rho$  exists, is it the unique solution satisfying (1.1)?

Note that question 2 above is as important as question 1: if such  $\rho$  existed but were not unique, there would be two different ways to assign page ranks to webpages and an arbitrary choice between the two would make many people very upset, if it changes the relative ranking of their webpage and the webpages of their business competitors.

It turns out that, in order to ensure that above two conditions are met, we must slightly alter equation (1.1) and refine our ranking model further. Most interestingly, besides ensuring the existence and uniqueness of the ranking function  $\rho$  which satisfies a reasonable modification of equation (1.1), such change comes naturally from another, entirely intuitive model or a “heuristics” for assigning page ranks, based on the concept of a *Random Surfer*, which we will describe later; at the moment, we will stick with such “imperfect” condition given by equation (1.1) and use it to introduce some necessary concepts and notation.

<sup>1</sup>The quotation marks indicate that we have not defined what we mean by “importance”; we are using this word in a rather vague manner.

As we have mentioned, since page ranks  $\rho(A_i)$  appear on both sides of equation (1.1), such equation cannot be used to actually compute page ranks because both  $\rho(P)$  and all of  $\rho(P_i)$  are initially unknown and must be assigned values simultaneously in a way such that (Formal) (1.1) is satisfied. Such conditions must be satisfied for *all* pages  $P \in \text{WWW}$  on the web. Thus, we have a system of equations, one for each page  $P$  on the web:

$$\left\{ \sum_{P_i \rightarrow P} \frac{\rho(P_i)}{\#(P_i)} \right\}_{P \in WWW} \quad (1.2)$$

The best way to write linear equations (linear in the values  $\rho(P_i)$ ) is using matrices.

Imagine now we write pages on the web at this time the matrix  $G_1$  corresponds and is given as  $g(i, j) = \frac{1}{\pi(P_i)}$  otherwise:



The  $M \times M$  where  $M = |WWW|$  is the total number of all web pages,  $P_1, P_2, \dots, P_M$  be an indexing of all such web pages. The  $i^{\text{th}}$  row of the entry in this row which is in column  $j$  is denoted by  $g(i, j)$  pointing to  $P_j$ , (in our notation, if  $P_i \rightarrow P_j$ ), and is equal to 0

WeChat: cstutorcs  
Assignment Project Exam Help  
Email: tutorcs@163.com

$$\text{QQ: } 749389476$$

$$g(i, j) = \begin{cases} \frac{1}{\binom{P_i}{j}} & \text{if } P_i \rightarrow P_j \\ 0 & \text{otherwise} \end{cases}$$

Thus,  $G_1$  consists mostly of zeros, because in the  $i^{th}$  row the non-zero entries are only in columns  $j$  such that  $P_i \rightarrow P_j$ , and each page has links to only a few other webpages, so  $G_1$  with its  $i^{th}$  row looks something like this:

where  $k$  is equal to  $\#(P_i)$ , the number of pages  $P_{j_1}, \dots, P_{j_k}$  which the page  $P_i$  has links to, and with non zero entries only in the corresponding columns  $j_1, \dots, j_k$ .

Recall from Linear Algebra that vectors  $\mathbf{r}$  are usually represented as column vectors; to make them into row vectors we have to transpose them. Thus, letting  $\mathbf{r}^T$  be the row vector of all page ranks:

$$\mathbf{r}^\top = (\rho(P_1), \rho(P_2), \dots, \rho(P_M))$$

the set of all equations (1.1) for all webpages  $\{P_1, \dots, P_M\}$  can now be written as a single matrix equation

$$\mathbf{r}^\top = \mathbf{r}^\top G_1 \quad (1.3)$$

This is because the  $j^{th}$  entry of the result of the multiplication on the right is obtained as the product of  $\mathbf{r}^T$  with the  $j^{th}$  column of  $G_1$ . Since the entries  $g(i, j)$  in the  $j^{th}$  column are non zero only for  $i_p$  such that  $P_{i_p} \rightarrow P_j$ , in which case  $g(i_p, j) = \frac{\rho(P_{i_p})}{\#(P_{i_p})}$ , we have that the  $j^{th}$  entry in the result of the multiplication  $\mathbf{r}^T G_1$  equals

$$(\rho(i), \rho(P_M)) \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{\#(P_{i_1})} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \frac{1}{\#(P_{i_p})} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{P_{i_p} \rightarrow P_j} \frac{\rho(P_{i_p})}{\#(P_{i_p})}$$

**WeChat: cstutorcs**  
**Assignment Project Exam Help**

Thus, the result of such multiplication would be equal to the  $j^{th}$  entry of  $\mathbf{r}^T$  just in case

**Email: tutorcs@163.com**

$$\rho(P_j) = \sum_{P_{i_p} \rightarrow P_j} \frac{\rho(P_{i_p})}{\#(P_{i_p})}$$

Notes:

**QQ: 749389476**

- Matrices of size  $M \times M$  which have much fewer than  $M^2$  non zero entries are called *sparse matrices*.
- We say that  $\lambda$  is a *left-hand eigenvalue* of a matrix  $G$  if there exist a vector  $\mathbf{x}$  such that  $\mathbf{x}^T G = \lambda \mathbf{x}^T$ ; vectors satisfying this property are called the (left-hand) eigenvectors corresponding to the eigenvalue  $\lambda$ .
- A matrix of size  $M \times M$  can have up to  $M$  distinct eigenvalues, which are, in general, complex numbers.
- $M\mathbf{x} = \lambda\mathbf{x}$  is equivalent to  $(M - \lambda I)\mathbf{x} = 0$ , where  $I$  is the identity matrix having 1 on the diagonal and zeros elsewhere. Such a homogeneous linear system has a non zero solution just in case the determinant of the system is zero, i.e.,  $\text{Det}(M - \lambda I) = 0$ , which produces a polynomial equation of degree  $n$  in  $\lambda$ ,  $p_n(\lambda) = 0$ .
- Polynomial  $p_n(\lambda)$  is called the characteristic polynomial of matrix  $M$ .
- Algebraic multiplicity of a solution  $\lambda_0$  for the equation  $p_n(\lambda) = 0$  is  $k$  just in case  $k$  is the largest power so that  $p_n(\lambda)$  is divisible by  $(\lambda - \lambda_0)^k$ .
- Geometric multiplicity of an eigenvalue  $\lambda_0$  is  $m$  just in case  $m$  is the largest number of linearly independent eigenvectors which correspond to that eigenvalue. It can be shown that the geometric multiplicity of an eigenvalue is smaller or equal than its algebraic multiplicity.
- An  $n \times n$  matrix is called *defective* if it does not have  $n$  linearly independent eigenvectors. Defective matrices must have at least one eigenvalue with algebraic multiplicity greater than 1.

Given that  $\mathbf{r}$  multiplies  $G_1$  from the left, the matrix equation (1.3), i.e.,  $\mathbf{r}^T = \mathbf{r}^T G_1$  simply says that:

- 1 is a left-hand eigenvalue of  $G_1$ ;
- $\mathbf{r}$  is a *left-hand eigenvector* of  $G_1$  corresponding to the eigenvalue 1.

Thus, finding ranks of the web pages would be reduced to finding eigenvectors of  $G_1$  which corresponds to the eigenvalue 1, providing that 1 is indeed one of the eigenvalues of  $G_1$ . But:

- Why should 1 be one of the eigenvalues of  $G_1$ ?

- Even if 1 is indeed an eigenvalue of  $G_1$ , there could be several corresponding linearly independent eigenvectors  $\mathbf{r}$ , and thus the page ranks which would be defined by such linearly independent vectors would be different (rather than one such ranking being just a scaled version of the others), and we saw that this could be bad.

## 程序代写代做 CS编程辅导

We now present another heuristic for ordering webpages by performing the following “thought experiment”.

Assume that someone starts at a web page on the web, and then clicks on a randomly chosen link on that web page; he continues to do so on each web page he visits, he randomly chooses a link on that page to get to the next page, and so on for a very, very large number of clicks, say  $T = 10^6 \times M$ , where  $M$  is the total number of pages on the web.

One could now argue that the probability of ending up on a web page  $P$  should equal to the ratio  $N(P)/T$  where  $N(P)$  is the number of times the surfer has visited the page  $P$  and  $T$  is the total number of clicks. This is because such a random surfer will frequently visit a webpage  $P$  just in case many other webpages, which themselves in turn are often pointed at, point to  $P$ .

However, there are immediate problems with such a proposal:

## WeChat: cstutorcs

1. What should such a surfer do if he gets to a web page without any outgoing links? Such pages are called “*dangling webpages*”. We can not ignore such webpages because they might be important, for example pdf documents. Having the surfer backtrack would involve keeping information about the previously visited pages which would tremendously complicate an analysis of the whole process.
2. Moreover, the statistical features of such web surfing should not depend on any particular choices of links the surfer makes; for as long as the number of clicks  $T$  is very large  $N(P)/T$  should be approximately the same for every particular surfing history. If we let  $T \rightarrow \infty$  the values  $N(P)/T$  should converge so that the page ranks eventually stay essentially the same, otherwise such page rank would be meaningless if it depended on either a particular surfing history or if it kept changing significantly as we increase the length of the surfing session.

## QQ: 749389476

One easy way to solve problem 1 is to have surfer jump to a randomly chosen webpage whenever he encounters a page without any outgoing links (a dangling webpage). As for the second problem, clearly in our present model the value of  $N(P)/T$  is not independent from particular surfing history of our surfer. Consider, for example, a set of webpages which form an isolated group, i.e., a set of webpages  $C$  which all have only links to other webpages from  $C$  but no outgoing links outside such a group and which are also all unknown to the rest of the web, in the sense that no webpage outside  $C$  points to a page in  $C$ . Then our random surfer will either never visit any webpage in the set  $C$  if his starting webpage is outside  $C$ , or will only visit pages in  $C$  if he starts from a webpage which is in  $C$ . Thus, in the first case all pages in  $C$  will get rank 0, while in the second case only webpages in  $C$  will have non zero page ranks.

One way to solve the second problem is to slightly change the behaviour of our surfer: every now and then, he gets bored following links on the pages he visits and decides to pick randomly a new webpage to start all over again.

Remarkably, these two modifications of the algorithm ensure that, given sufficiently large  $T$ , for any webpage  $P$  the value of  $N(P)/T$  will be essentially independent of both the particular surfing history (i.e., independent on which particular outgoing links he chooses to follow on any particular webpage) and independent of the duration of the surfing, i.e., the values of  $N(P)/T$  converge to a value which we can call the PageRank. More over, as we will see, these two fixes also allow a computation of a good approximation of such values computable without actually performing such long web surfing.

We now present another, closely related heuristics of how to assign the page rank.

Let us consider an arbitrary starting webpage  $P$  and all possible surfing histories of length  $T$ . For every other webpage  $P_i$  on the Internet we could count how many surfing histories of length  $T$  end on that webpage  $P_i$ , and divide it with the total number of surfing histories of length  $T$  to get a ratio  $\rho_i$ . This ratio would give us the probability that, staring from page  $P$  and after a surfing session of  $T$  many clicks, we end up on page  $P_i$ . Can we consider such a probability a good measure of importance of each page  $P_i$ ? For this to be true such  $\rho_i$  should be “almost” independent of the starting page and independent of  $T$  for as long as  $T$  is sufficiently large. By this we mean that for any two webpages  $P_1$  and  $P_2$  and integers  $T_1$  and  $T_2$ , surfing sessions starting from page  $P_1$  (page  $P_2$  respectively) and following  $T_1$  ( $T_2$ , respectively) many links, the corresponding ratios  $\rho_1^1$  and  $\rho_2^2$  for any webpage  $P_i$  should be very close in value. In more formal terms, as  $T \rightarrow \infty$  the probabilities  $\rho_i$  should converge, and should converge to numbers which are independent of the starting point.

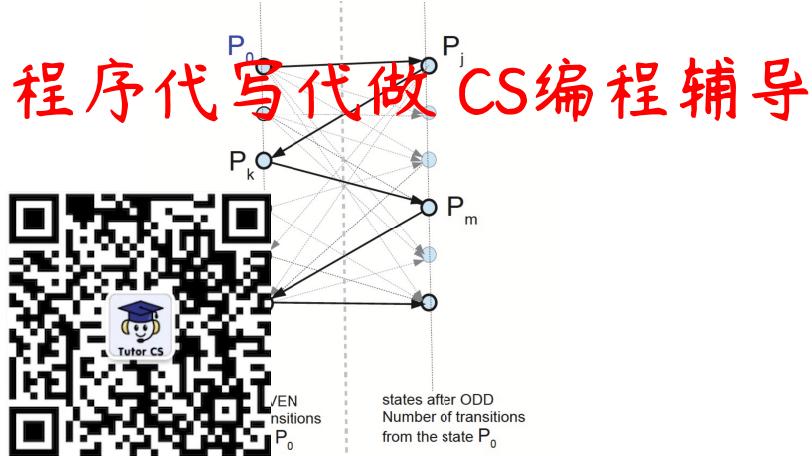


Figure 1.1: A “bipartite internet”  
**WeChat: cstutorcs**

However, such a property can fail if the underlying directed graph of webpages has nodes with “periodicity”, for example, if the graph was bipartite; see Figure 1. In such a case if the surfer starts from the left side of the graph, the probability for him to be still on the left side after any odd number of clicks would be zero. Thus, probability to be found at a particular node after  $T$  many clicks would not converge as  $T$  gets large. However, the modified surfer behaviour we introduced above, in which the surfer occasionally interrupts following links on page he visits and restarts from a new, arbitrarily chosen page, in fact guarantees that  $n(P)/T$  will converge to a well defined probability of finding a surfer at a particular page  $P$  after a surfing session of  $T$  many clicks.

**Email: tutorcs@163.com**

We now wish to produce a matrix representation of such modified surfing session. Consider again the **initial**, unmodified matrix  $G_1$ :

$$G_1 = \begin{pmatrix} \text{QQ: 749389476} & & & \\ & \vdots & \vdots & \vdots \\ & \text{https://tutorcs.com} & & \vdots \\ & \dots 0 0 0 0 0 \dots & \dots 0 0 0 0 0 \dots & \dots 0 0 0 0 0 \dots \\ & \dots 0 0 \frac{1}{\#(P_i)} 0 0 \dots & \dots 0 0 \frac{1}{\#(P_i)} 0 0 \dots & \dots 0 0 \frac{1}{\#(P_i)} 0 0 \dots \\ & \vdots & \vdots & \vdots \\ & \vdots & \vdots & \vdots \end{pmatrix}$$

The  $i^{th}$  row of such a matrix corresponds to the  $i^{th}$  web page and is either a row of zeros if  $P_i$  is a dangling page without any outgoing links, or a sparse vector with non zero entries only at columns corresponding to pages  $P_j$  such that  $P_i \rightarrow P_j$ , in which case the value is equal  $1/\#(P_i)$ , i.e., the reciprocal of the total number  $\#(P_i)$  of outgoing links that  $P_i$  has.

This matrix has an interpretation that the entries in each row are *the probabilities* that if the surfer is visiting page  $P_i$  he will visit page  $P_j$  next, because he chooses with equal probability any link on  $P_i$ ; thus, the probability to go next to the page  $P_j$  is nonzero only if  $P_i$  has a link to the page  $P_j$ , in which case such probability is equal for all such pages,  $1/\#(P_i)$ .

What does  $G_2$  look like if it is obtained from  $G_1$  by fixing the problem of dangling webpages? If a page  $P_i$  is dangling, then all entries in row  $i$  of  $G_1$  are 0, and in such case we decided to have our surfer jump to a random new page. Thus, in the new matrix  $G_2$  of probabilities of transitions from a page  $P_i$  to a page  $P_j$  the rows corresponding to all dangling pages have values all equal to  $1/M$  instead of 0, because all webpages are equally likely to be the destinations from a dangling page, and since there are  $M$  webpages on the web in total, each page is picked with probability  $1/M$ . Thus we get a new matrix  $G_2$  which looks as follows:

$$G_2 = \left( \begin{array}{cccccc|ccccc} \vdots & \vdots \\ \cdots \frac{1}{M} & \text{QR code} & \cdots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots & \cdots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots \\ \cdots 0 & 0 & \frac{1}{\#(P_i)} & 0 & 0 \cdots & \cdots 0 & 0 & \frac{1}{\#(P_i)} & 0 & 0 \cdots \\ \vdots & \vdots \end{array} \right)$$

Note that in the above matrix each row sums up to 1. Such matrices are called **row stochastic** and, as we will see, they have some really useful properties.

**WeChat: cstutorcs**

To fix the second problem and prevent the surfer to become trapped for ever in a set of webpages without links going out of such a set, we have decided to have him either follow an arbitrary link on a webpage he visits with a probability  $\alpha$ , or, with probability  $1 - \alpha$  jump to an arbitrarily chosen webpage from the entire web. Thus we modify further matrix  $G_2$  to obtain the *Google Matrix*  $G$  defined as

**Assignment Project Exam Help**

$$G = \left( \begin{array}{cccccc|ccccc} \vdots & \vdots \\ \cdots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots & \cdots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots \\ \cdots \frac{1-\alpha}{M} & \frac{1-\alpha}{M} & \frac{\alpha}{\#(P_i)} + \frac{1-\alpha}{M} \frac{1-\alpha}{M} \cdots & \cdots \frac{1-\alpha}{M} & \frac{1-\alpha}{M} & \frac{\alpha}{\#(P_i)} + \frac{1-\alpha}{M} & \frac{1-\alpha}{M} & \frac{1-\alpha}{M} & \frac{1-\alpha}{M} \cdots \\ \vdots & \vdots \\ \text{Email: tutorcs@163.com} & \text{QQ: 749389476} & \vdots \end{array} \right)$$

<https://tutorcs.com>

The last transformation does not change the rows corresponding to dangling webpages because at such webpages we go to an arbitrary page with probability  $\alpha/M + (1 - \alpha)/M = 1/M$ .

The first fix was obtained by starting with the original matrix  $G_1$  and adding to it a matrix which has non-zero entries only in the rows corresponding to the dangling webpages, where it always has value  $1/M$ :

$$G_2 = G_1 + \left( \begin{array}{cccccc|ccccc} 0 & \dots 0 & 0 & 0 & \dots & \dots 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & \vdots & & & & \vdots \\ 0 & \dots 0 & 0 & 0 & \dots & \dots 0 & 0 & 0 & \dots & 0 \\ \frac{1}{M} & \dots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots & \cdots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots & \cdots \frac{1}{M} & \frac{1}{M} \cdots & \frac{1}{M} \\ 0 & \dots 0 & 0 & 0 & \dots & \dots 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & \vdots & & & & \vdots \\ 0 & \dots 0 & 0 & 0 & \dots & \dots 0 & 0 & 0 & \dots & 0 \\ \frac{1}{M} & \dots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots & \cdots \frac{1}{M} & \frac{1}{M} & \frac{1}{M} \cdots & \cdots \frac{1}{M} & \frac{1}{M} \cdots & \frac{1}{M} \\ 0 & \dots 0 & 0 & 0 & \dots & \dots 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & \vdots & & & & \vdots \\ 0 & \dots 0 & 0 & 0 & \dots & \dots 0 & 0 & 0 & \dots & 0 \end{array} \right)$$

Let us consider vectors<sup>2</sup>  $\mathbf{d}, \mathbf{e} \in \mathbb{R}^M$  as follows: vector

$$\mathbf{d}^\top = (0 \dots 0 \ 1 \ 0 \dots 0 \ 1 \ 0 \dots 0 \ 1 \ 0 \dots 0)$$

is one at positions  $i$  which correspond to dangling webpages and zero elsewhere; vector

$$\mathbf{e}^\top = (1 \ 1 \ \dots \ 1)$$

---

<sup>2</sup>Vectors are represented as columns; thus, to make them into row vectors we have to transpose them.

is one everywhere. We now have

$$G_2 = G_1 + \frac{1}{M} \mathbf{d} \mathbf{e}^T$$

# 程序代写代做 CS 编程辅导

Note that the product of a column vector  $\mathbf{d}$  with a row vector  $\mathbf{e}^T$  (both of length  $M$ ) is a matrix of size  $M \times M$  with non-zero rows corresponding to the indices of dangling webpages  $P_i$ .

We get the final matrix  $G$  by making a convex combination of  $G_2$  and a matrix whose all entries are  $1/M$ , i.e.,



$$= \alpha \left( G_1 + \frac{1}{M} \mathbf{d} \mathbf{e}^T \right) + \frac{1-\alpha}{M} \mathbf{e} \mathbf{e}^T, \quad (1.4)$$

where  $0 < \alpha < 1$  is a fixed probability that the surfer will follow a link on a webpage he visits; thus,  $1 - \alpha$  is the probability instead jump to a randomly chosen webpage. Note that  $G$  is still row stochastic because the matrix  $1/M \mathbf{e} \mathbf{e}^T$  are row stochastic, and thus so is their convex combination. Note also that if you start with a vector whose coordinates sum up to 1 with a row stochastic matrix you get a vector whose coordinates again sum up to 1.

While  $G$  defined above is no longer sparse, the product of any vector  $\mathbf{x}^T$  and the matrix  $G$  can be decomposed into a sum:

$$\begin{aligned} \mathbf{x}^T G &= \alpha \left( \mathbf{x}^T G_1 + \frac{1}{M} \mathbf{x}^T \mathbf{d} \mathbf{e}^T \right) + \frac{1-\alpha}{M} \mathbf{x}^T \mathbf{e} \mathbf{e}^T \\ &= \alpha \mathbf{x}^T G_1 + \frac{1}{M} (\alpha \mathbf{x}^T \mathbf{d} + (1-\alpha) \mathbf{x}^T \mathbf{e}) \mathbf{e}^T \\ &= \alpha \mathbf{x}^T G_1 + \frac{1}{M} (\alpha \mathbf{x}^T \mathbf{d} + 1 - \alpha) \mathbf{e}^T \\ &= \alpha \mathbf{x}^T G_1 + \frac{1}{M} (1 - \alpha(1 - \mathbf{x}^T \mathbf{d})) \mathbf{e}^T \end{aligned}$$

# WeChat: cstutorcs

# Assignment Project Exam Help

# Email: tutorcs@163.com

QQ: 749389476

with the third equality holding because  $\mathbf{x}^T \mathbf{e} = 1$ , i.e., because all coordinates of  $\mathbf{x}$  sum up to 1 (recall that the coordinates of  $\mathbf{x}$  are interpreted as probabilities of visiting webpages). Note that  $\mathbf{x}^T \mathbf{d}$  is just the probability to be at a dangling webpage. Now each summand can be computed efficiently and only the original matrix  $G_1$  and vector  $\mathbf{d}$  need to be stored.

The reason why our modified surfer heuristics, with probabilities of transitioning from a webpage to another webpage given by the above matrix  $G$ , allows a correct and unique PageRank assignment is a very general one: our random surfer is a special instance of a *Markov Chain* (also called a *Discrete Time Markov Process*).

## 2 Markov Chains

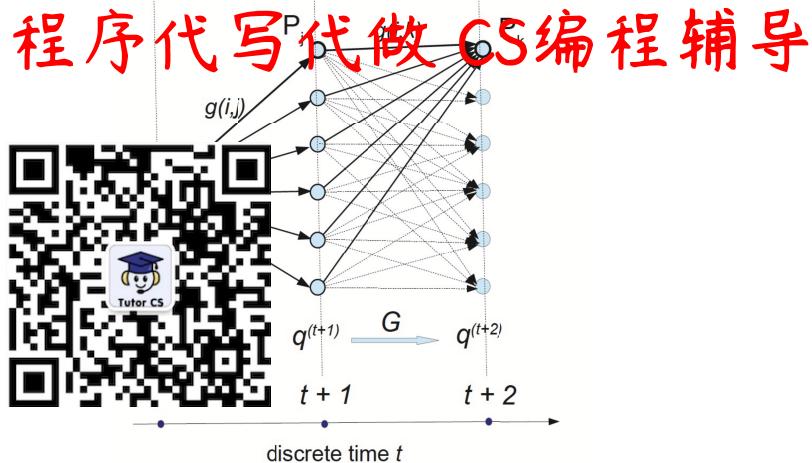
A (finite) **Markov Chain** is given by a finite set of states  $S = \{P_i\}_{i \leq M}$  and by a row stochastic matrix  $G$ . The process can start at  $t = 0$  in any of its states and continue its evolution by going at every discrete moment of time  $t$  (i.e., at any integer  $t$ ) from its present state  $X(t) = P_i$  into another, randomly chosen state  $X(t+1) = P_j \in S$ , such that, if the present state is  $P_i$ , then the probability for the system entering a state  $P_j$  at the next instant is given by the entry  $g(i, j)$  in the  $i^{th}$  row and  $j^{th}$  column of the matrix  $G$ . Thus, such probability DOES NOT depend on the way the state  $P_i$  was reached (i.e., it does not depend on the history of the process) and does not depend on the particular moment of time when such a transition happens; it only depends on what the present state of the system is.

Our random surfer's behaviour is an example of a Markov Chain, with all the web pages on the internet being the states of the system, and with the matrix  $G$  as the matrix of probabilities of transition from one state into another, i.e., from one webpage to the next one. Thus, if the surfer is at the page  $P_i$ , entry  $g(i, j)$  is the probability that the surfer will be at the next instant of time at a page  $P_j$ , either by:

- following a link on the page  $P_i$  which points to page  $P_j$  (if such a link exists) with probability  $\frac{\alpha}{\#(P_i)}$ , or
- by jumping directly to page  $P_j$ , if page  $P_j$  happens to be the one chosen at random, with probability  $\frac{1-\alpha}{M}$ .

Let  $q_i^{(t)}$  be the probability that the state at a moment  $t$  be  $P_i$ , i.e.,  $X(t) = P_i$  and let  $(\mathbf{q}^{(t)})^T = (q_1^{(t)}, \dots, q_i^{(t)}, \dots, q_M^{(t)})$  be the corresponding row vector.

If at the instant  $t$  the system is in state  $P_i$ , then at the instant  $t+1$  the system will be in a state  $P_j$  precisely with probability  $g(i, j)$  because this is the definition of the matrix  $G$  associated with a Markov chain. What is the probability that at the next instant  $t+2$  the system will be in a state  $P_k$ ? We simply sum all probabilities



## WeChat: cstutorcs

Figure 2.1: Evolution of Markov chain probabilities of states

of getting into state the  $P_j$  through all possible intermediate states  $P_j$  at the instant  $t + 1$  i.e., we sum the probabilities of the system being at  $t + 1$  in any possible state times the probability of transitioning from such intermediary state into the state  $P_k$ :

$$q_k^{(t+2)} = \sum_{1 \leq j \leq M} q_j^{(t+1)} g(j, k)$$

**Email: tutorcs@163.com**

(see figure 2.1) which simply says that

$$(\mathbf{q}^{(t+2)})^\top = (\mathbf{q}^{(t+1)})^\top G$$

We can now apply the very same reasoning to obtain the probability distribution  $\mathbf{q}^{(t+3)}$  for the states the system can be in at instant  $t + 3$ , by simply applying the matrix  $G$  to  $\mathbf{q}^{(t+2)}$  from the right and obtain

$$(\mathbf{q}^{(t+3)})^\top = (\mathbf{q}^{(t+2)})^\top G = (\mathbf{q}^{(t+1)})^\top G G = (\mathbf{q}^{(t+1)})^\top G^2$$

In fact, rather than assuming that the system started from any particular state, we can assume that we have a probability distribution  $\mathbf{q}^{(0)}$  for possible initial states at the instant  $t = 0$  and obtain that in this case the probability distribution of the states at an instant  $k$  is given by

$$\mathbf{q}^{(k)} = \mathbf{q}^{(0)} G^k$$

### 2.1 Graphs associated with non-negative matrices

To any non negative matrix  $M \geq 0$ , i.e., a matrix such that every entry  $M_{ij}$  satisfies  $M_{ij} \geq 0$ , if  $M$  is of size  $n \times n$  we can associate an underlying directed graph with  $n$  vertices  $V = \{P_1, P_2, \dots, P_n\}$  and a directed edge  $e(i, j)$  just in case  $M_{ij} > 0$ . The adjacency matrix  $\tilde{M}$  of such a graph is obtained from the matrix  $M$  by setting all non zero entries to 1. It is now easy to verify that there is a path of length  $k$  from  $P_i$  to  $P_j$  just in case  $(\tilde{M}^k)_{ij} > 0$ .

**Homework Problem 1.** Prove by induction on  $k$  that  $(\tilde{M}^k)_{ij}$  is exactly the number of directed paths from  $P_i$  to  $P_j$  of length exactly  $k$ .

**Definition:** We say that a directed graph  $G$  is *strongly connected* just in case for any two vertices  $P_i, P_j \in G$  there exists a directed path in  $G$  from  $P_i$  to  $P_j$  (and thus also a directed path from  $P_j$  to  $P_i$ ).

**Definition:** A Markov chain is *irreducible* if the graph corresponding to its transition probabilities matrix is strongly connected.

In our example of a random surfer, the irreducibility of the underlying graph entails that there are no groups of web pages which are traps from which the surfer cannot escape if he happens to visit one of these webpages. We guarantee that there cannot be such traps by ensuring that every once in a while (with probability  $1 - \alpha$ , to be precise), our surfer jumps to a randomly picked webpage from the entire web to continue his surfing. Thus, every two vertices in the underlying graph are connected, possibly with a very low transition probability  $(1 - \alpha)/M$  (recall that  $M$  is the total number of webpages on internet).

We would now like to have the property that for every initial probability distribution  $\mathbf{q}^{(0)}$  the values of  $\mathbf{q}^{(k)}$  converge to some probability distribution  $\mathbf{q}$  independent of  $\mathbf{q}^{(0)}$ . However we saw that if the underlying graph  $G$  is bipartite and  $q_i^0 = 0$  for all  $i$  except for some  $i_0$  for which  $q_{i_0}^0 = 1$ , then, regardless of the same side of such partition (see again figure 1) as  $S_{i_0}$  such value after odd many transitions, i.e.,  $q_{i_0}^{(k+1)}$ , is 0. Thus, such values cannot converge. It turns out that such periodicity (like the one in the case of a bipartite graph) is the only leftover cause why  $\mathbf{q}^{(t)}$  might fail to converge.

**Definition:** A state  $P_i$  is

periodic if there exists an integer  $K > 1$  such that all loops in its underlying graph which contain  $P_i$  have length divisible by  $K$ . Markov chains which do not have any periodic states are called aperiodic.

The following theorem (known as Google page rank theorem) is a very important and general property of Markov chains which insures that the Google page rank is well defined (unique) and that it can be computed by an iterative procedure.

**Theorem 2.1** Any finite, irreducible and aperiodic Markov chain has the following properties:

1. For every initial probability distribution of states  $\mathbf{q}^{(0)}$  the value of  $(\mathbf{q}^{(t)})^\top = (\mathbf{q}^{(0)})^\top G^t$  converges as  $t \rightarrow \infty$  to a unique stationary distribution  $\mathbf{q}$ , i.e., converges to a unique distribution  $\mathbf{q}$  which is independent of  $\mathbf{q}^{(0)}$  and satisfies  $\mathbf{q}^\top = \mathbf{q}^\top G$ .
2. Let  $N(P_i, T)$  be the number of times the system has been in state  $P_i$  during  $T$  many transitions of such a Markov chain; then

**Assignment Project Exam Help**

$$\lim_{T \rightarrow \infty} \frac{N(P_i, T)}{T} = \mu_i$$

Note that the above theorem provides precisely what we need. First of all, the theorem applies to the Google matrix  $G$ : by taking care of the dangling webpages the initial hyperlink matrix  $G$  has been first turned into a stochastic matrix  $G_2$ , which allows us to consider its entries as transition probabilities; the second fix with allowing the surfer to “teleport” at any instant with a small probability  $1 - \alpha$  to a new, randomly chosen webpage precludes both the existence of periodic states, because one can get with a small, but non-zero probability of  $(1 - \alpha)/M$  to any page at any instant  $t$ , as well as makes the underlying graph strongly connected. Consequently, the process is irreducible and aperiodic and the above theorem on Markov chains implies that:

- 1 is a left eigenvalue of the Google matrix  $G$ , and the stationary distribution  $\mathbf{q}$  is the corresponding left hand side eigenvector,  $\mathbf{q}^\top = \mathbf{q}^\top G$ ;
- such stationary distribution  $\mathbf{q}$  is unique, i.e., if  $\mathbf{q}^\top = \mathbf{q}^\top G$  and  $\mathbf{q}_1^\top = \mathbf{q}_1^\top G$ , then  $\mathbf{q}_1 = \mathbf{q}$ ;
- distribution  $\mathbf{q}$  can be obtained by starting with an arbitrary initial probability distribution  $\mathbf{q}_0$  and obtain  $\mathbf{q}$  as  $\lim_{k \rightarrow \infty} \mathbf{q}_0^\top G^k$ ;
- an approximation  $\tilde{\mathbf{q}}$  of  $\mathbf{q}$  can be obtained by taking  $\mathbf{q}_0^\top = (1/M, 1/M, \dots, 1/M)$  and a sufficiently large  $K$  and computing  $\tilde{\mathbf{q}} = \mathbf{q}_0^\top G^K$  iteratively (NOT by computing  $G^K$ ) via:

$$\begin{aligned} \mathbf{q}(0) &= \mathbf{q}_0 \\ (\mathbf{q}(n+1))^\top &= (\mathbf{q}(n))^\top G \quad \text{for } 0 \leq n < K; \end{aligned} \tag{2.1}$$

- the  $i^{th}$  coordinate of such obtained distribution  $\mathbf{q}^\top = (q_1, \dots, q_i, \dots, q_M)$  roughly gives the ratio  $N(P_i, T)/T$  where  $N(P_i, T)$  is the number of times the surfer has visited page  $P_i$  during a surfing session which lasted  $T$  many clicks, because if  $T$  is very large, the probability of being at a particular webpage  $P_i$  will be very close to  $q_i$  for the most of the surfing history, and thus the number  $N(P_i, T)$  of visits of a page  $P_i$  will be roughly equal to  $q_i T$ .

How close to 1 should  $\alpha$  be?? Let us compute the *expected* length  $lh$  of surfing between two “teleportations” to a randomly chosen website, i.e., of surfing by only following links, ignoring the dangling webpages where the random surfer must also teleport to a new, randomly chosen location.

Assume that the surfer has just landed at a new, randomly chosen website. The probability that  $lh = 0$ , i.e., to immediately teleport is  $1 - \alpha$ ; the probability to follow one link and then teleport is  $\alpha(1 - \alpha)$ ; similarly, to follow  $k$  links and then teleport is  $\alpha^k(1 - \alpha)$ . Thus, the expected duration of the surfing is just the expectation

of the *geometric probability distribution*, also called the probability distribution of the number of Bernouli trials; the probability of the random variable to take value  $k$  is equal to  $\alpha^k(1 - \alpha)$ ; thus,

$$\begin{aligned} E(lh) &= 0(1 - \alpha) + 1\alpha(1 - \alpha) + \dots + k\alpha^k(1 - \alpha) + \dots \\ &= \alpha(1 - \alpha)(1 + 2\alpha + \dots + k\alpha^{k-1} + \dots) \end{aligned}$$

We now evaluate the sum



ice trick: let us set

$$s = 1 + 2\alpha + 3\alpha^2 + \dots + k\alpha^{k-1} + \dots;$$

then

$$\begin{aligned} s &= 1 + 2\alpha + \frac{1}{1-\alpha}(1 + 2\alpha + 3\alpha^2 + \dots + k\alpha^{k-1} + \dots) \\ &= (1 + \alpha + \alpha^2 + \alpha^3 + \dots + \alpha^{k-1} + \dots) + (\alpha + 2\alpha^2 + 3\alpha^3 + \dots + (k-1)\alpha^{k-1} + \dots) \\ &= \frac{1}{1-\alpha} + \alpha(1 + 2\alpha + 3\alpha^2 + \dots + (k-1)\alpha^{k-2} + \dots) \\ &= \frac{1}{1-\alpha} + \alpha s; \end{aligned}$$

so to find  $s$  it suffices to solve the equation  $s = \frac{1}{1-\alpha} + \alpha s$  which gives  $s = 1/(1-\alpha)^2$ . Thus,

**Assignment Project Exam Help**

$$E(lh) = \alpha(1 - \alpha) \frac{1}{(1 - \alpha)^2} = \frac{\alpha}{1 - \alpha}$$

Google uses  $\alpha = .85$ ; thus the expected surfing length is  $\frac{.85}{1 - .85} = \frac{.85}{.15} \approx 5.7$ . Consequently, on average, less than 6 links are followed between two teleportations! Is that enough to determine which pages are important? The answer is given by the extraordinary success of Google: even though it looks small, it must obviously be just enough. This is not so surprising: if a page is endorsed indirectly by an important webpage through several webpages as intermediaries, in all likelihood this chain is less than 6 links long. You might want to read about *six degrees of separation*, see Chapter 9 in the textbook “Networked Life” or

[http://en.wikipedia.org/wiki/Six\\_degrees\\_of\\_separation](http://en.wikipedia.org/wiki/Six_degrees_of_separation).

Google inventors Page and Brin obtained the value  $.85$  empirically: for such a value of  $\alpha$  the number of iterations needed to obtain a reasonably good approximation of the page rank is between  $50 - 100$ , i.e.,  $\mathbf{q} \approx \mathbf{q}_0 G^{50}$ . Larger values seem to produce more accurate representation of “importance” of a webpage, but the convergence slows down fast and thus the number of iterations  $K$  in (2.1) must be taken significantly greater than 100. In fact, as we will see later, the rate of convergence of the so called “power method”, i.e., iterative computation of  $\mathbf{q} \approx \mathbf{q}_0 G^k$  depends on the ratio of the second largest eigenvalue and the largest eigenvalue of  $G$ . We will show later that 1 is the largest eigenvalue of  $G$  and that the second largest eigenvalue is  $\approx \alpha$ . Consequently, the error of approximation of  $\mathbf{q}$  by  $\mathbf{q}_0 G^k$  decreases approximately as  $\alpha^k$ . Thus, to compare the necessary number of iterations for  $.85$  and  $.95$  we see that if  $.95^m = .85^k$ , then, by taking the logarithm of both sides,  $m \log_{10} .95 = k \log_{10} .85$ , i.e.,  $m = k \log_{10} .85 / \log_{10} .95 > 3.168k$ , i.e., the number of iterations with  $\alpha = .95$  is more than three times larger than the number of iterations necessary with  $\alpha = .85$ , if one is to achieve the same number of significant digits.

Also, increasing  $\alpha$  increases the sensitivity of the resulting PageRank, in the sense that small perturbations of the link structure of the internet can change the page rank considerably, which is a bad thing: if Google ranking criteria changed from day to day, such a rank would be of a dubious value.

The performance of algorithms “for human use” (such as search algorithms) often crucially depends on the choice of the parameters involved; while analytic estimates of the behaviour of the algorithm such as the sensitivity to perturbations of values of its inputs or changes of the parameters are valuable and in fact often necessary pointers to know how to *initially* chose these parameters, only empirically assessed “real life” performance provides a way to properly fine-tune these parameters.

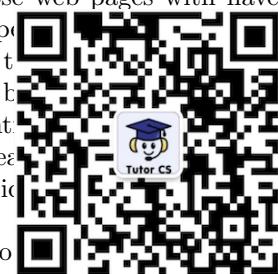
## 2.2 Refinements of the PageRank algorithm

While the above algorithm apparently still represents the core of Google’s search technology, the algorithm had been improved and “tweaked” in many ways.<sup>1</sup> While the details are closely guarded Google’s proprietary

<sup>1</sup>Some allege that it has also been altered in unfair ways, making it favour results which suit Google business model, centred on advertising.

information, some possible improvements are pretty straightforward to come up with.

First of all, one can argue that not all outgoing links from a web page are equally important. For example, a link to a legal disclaimer is not something people would often search for on a web page; thus, it should arguably get much lower weight than  $1/\#(P)$ . One could, for example, argue that one should give higher weight to links which point to those web pages with have similar kind of content as the web page containing the link, but this is a slippery slope. The notion which is notoriously difficult to capture - *the semantics* i.e., the *meaning* rather than the *strings of symbols*). The best way to simply count how many times visitors chose to follow a particular link and give each link a probability equal to the ratio of the number of times visitors chose to follow that particular link to the total number of clicks on links on that page.



Secondly, one can also argue that “uniform” components of  $G$ , i.e., matrices  $\frac{\alpha}{M} \mathbf{d} \mathbf{e}^T$  and  $\frac{1-\alpha}{M} \mathbf{e} \mathbf{e}^T$ , are unrealistic: if one is searching for a topic about sports, he will not chose to stop following links on the webpage he is at, and, by a truly random choice, pick as his next web page to visit the web page with the list of staff of CSE. Similarly, if he ended up at a “dangling webpage” related to sports, his next, randomly chose webpage should also be sports related.

One can, in fact, in the equation (1.4),

$$G = \alpha \left( G_1 + \frac{1}{M} \mathbf{d} \mathbf{e}^T \right) + \frac{1-\alpha}{M} \mathbf{e} \mathbf{e}^T \quad (2.2)$$

instead of using  $\frac{\alpha}{M} \mathbf{d} \mathbf{e}^T$  and  $\frac{1-\alpha}{M} \mathbf{e} \mathbf{e}^T$  use a topic specific teleportation matrix of the form  $\alpha \mathbf{d} \mathbf{v}^T$  and  $(1-\alpha) \mathbf{e} \mathbf{v}^T$  where  $\mathbf{v}$  is still a probability distribution (i.e., all coordinates sum up to 1), whose all coordinates are non zero (to keep the corresponding Markov process irreducible and aperiodic), but where coordinates  $v_i$  of  $\mathbf{v}$  which correspond to web pages which are relevant to that particular topic are of significantly larger value than for those pages which are not relevant for this topic. Note that in this way the resulting correction is still representable in an equally compact way and allows almost equally fast computation of the corresponding topic specific page rank as the “uniform” matrices  $\frac{\alpha}{M} \mathbf{d} \mathbf{e}^T$  and  $\frac{1-\alpha}{M} \mathbf{e} \mathbf{e}^T$ .

There are several problems with implementing the above idea. Each different topic vector  $\mathbf{v}$  would produce a different page rank which would have to be separately computed using the iterative algorithm we described. Since such computation requires very large computational power and time, it must be done in advance for each of the very large number of topics of interests across entire web community, which is clearly not feasible, since a single computation (at least 1000) last several days.

A solution was offered by another Stanford University startup, *Kaltix*, acquired by Google only a few months after its inception. The solution is based on the following idea: all web pages are very broadly classified (by, say, the key words they contain) into a small number  $K$  of extremely general categories corresponding to “topics”; (*Kaltix* started with only 16 categories). Each of these categories has a corresponding “personalization” vector  $\mathbf{v}(k)$ ,  $1 \leq k \leq K$ , as described before, and the corresponding PageRank vectors  $\rho_k$  are computed.

For each query one can now use some separate algorithm to decide to what extent each of the  $K$  topics is relevant. For example, searching for “Google lawsuits” would probably give to topics such as “law and policy” high weight, as well as high weights to topics such as “technology” and “business”, but zero weight to topics such as “sport” or “food”. One can now obtain a very “reasonably personalised” page ranks of all web pages simply by computing a linear combination of the page ranks  $\rho_i$  corresponding to each of  $K$  many topics, i.e., of the form

$$\rho^*(P) = \sum_{1 \leq j \leq K} \frac{q_j}{\sum_{1 \leq i \leq K} q_i} \rho_j$$

where  $q_i$  is the weight given to  $i^{th}$  topic; the denominator is used to normalize the weights so that they sum up to 1, thus guaranteeing that the obtained linear combination is a correct form of a weighted average.

Of course, Google algorithm must have been (and continues to be) ‘tweaked’ in may other ways, for example to defeat “link farms” and other web sites which are trying to increase their page rank without real merit, but the bottom line is this: it was an ingenious idea how to make web ranking much more resilient to manipulation, by making such evaluation “global”, in the sense that the page rank of each web page depends on the ranks of ALL other web pages on the web. In this way “localized” attacks, even if they involve quite a few other web pages are much less likely to succeed. While both the Markov chains and random walks on graphs have been studied “ad nauseam” and are thus far from being a novelty, the Google inventors deserve a huge credit for finding the ultimate present day application of these “ancient” concepts.

**Homework Problem 2.** The present day “publish or perish” madness in academia involves counting number of papers researchers have published, as well as the number of citations their papers got.

- One might argue that not all citations are equally valuable: a citation in a paper that is itself often cited is more valuable than a citation in a paper that no one cites. Design a PageRank style algorithm which would rank papers according to their “importance” and then use such an algorithm to rank researchers by their “importance”.
- Assume now that you do not have information on the citations in each published paper, but instead you have for every researcher a list of all other researchers who have cited him and how many times they cited him. Design again a PageRank style algorithm which would rank researchers by their importance.
- What do you think, which of the two methods of ranking researchers would do more damage to science?



## Related Material

We now explore the PageRank in greater depth; we start by presenting (a part of) the Perron Frobenius theory for non-negative matrices and its application to PageRank.

### 3 A brief introduction to the Perron-Frobenius theory

We start with a few definitions.

#### Assignment Project Exam Help

- A real matrix  $M$  is **non-negative (positive)** if all of its entries satisfy  $M_{ij} \geq 0$  ( $M_{ij} > 0$ ); we denote this fact by  $M \geq 0$  ( $M > 0$ , respectively). Accordingly, if  $\mathbf{u}$  is a vector, then  $\mathbf{u} \geq 0$  means that every coordinate  $u_i$  of  $\mathbf{u}$  is **non-negative**,  $\mathbf{u} > 0$  means that every coordinate is **strictly positive**.
- A non-negative square matrix  $M$  of size  $n \times n$  is **irreducible** if for every  $1 \leq i, j \leq n$  there exists a natural number  $k \geq 1$  such that  $(M^k)_{ij} > 0$ .
- A non-negative square matrix  $M$  is **primitive** if there exists a natural number  $k$  such that  $M^k$  is positive.
- The **spectral radius**  $\rho(M)$  of a square matrix  $M$  is defined as

$$\rho(M) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } M\}.$$

<https://tutorcs.com>

Note that, if a matrix  $P > 0$  is positive and  $\mathbf{u} \leq \mathbf{v}$ ,  $\mathbf{u} \neq \mathbf{v}$ , then  $0 \leq \mathbf{v} - \mathbf{u} \neq 0$  which implies  $0 < P(\mathbf{v} - \mathbf{u})$  and thus  $P\mathbf{u} < P\mathbf{v}$ .

We now present elements of the Perron-Frobenius theory for non-negative matrices; rather than developing it in full generality, we do what is just enough for our purpose.

**Theorem 3.2** Let  $M$  be an  $n \times n$  irreducible matrix; then:

- $M$  has a real positive eigenvalue  $\lambda_{\max} > 0$  such that every other eigenvalue  $\lambda$  satisfies  $|\lambda| \leq \lambda_{\max}$ .
- The eigenvectors  $\mathbf{x}$  corresponding to  $\lambda_{\max}$  can be chosen so that  $\mathbf{x} > 0$ , and if an eigenvector  $\mathbf{y}$  satisfies  $\mathbf{y} \geq 0$ , then it must correspond to  $\lambda_{\max}$ .

**Proof:** Let  $V^+ = \{\mathbf{y} : \mathbf{y} \geq 0 \& \mathbf{y} \neq 0\}$ ; we define a function  $f$  on  $V^+$  by

$$f(\mathbf{y}) = \min \left\{ \frac{(M\mathbf{y})_i}{y_i} : 1 \leq i \leq n, y_i \neq 0 \right\};$$

clearly,  $f$  satisfies  $f(\mathbf{y}) = \max\{t : t\mathbf{y} \leq M\mathbf{y}\}$ .

Let  $S^+$  be the section of the unit sphere with all coordinates non negative, i.e.,  $S^+ = \{\mathbf{y} : \|\mathbf{y}\| = 1 \& y_i \geq 0\}$ . Note that, by linearity of  $M$ , for every positive constant  $c > 0$ ,  $f(c\mathbf{y}) = f(\mathbf{y})$ ; thus, all values which  $f$  attains on  $V^+$ ,  $f$  also attains on just  $S^+$ , which is a compact set.

Note that if  $M$  is irreducible, then matrix  $I + M$  is primitive, because

$$(I + M)^k = \sum_{i=0}^k \binom{k}{i} M^i \quad (3.1)$$

and so, if  $k$  is large enough, eventually all entries in  $(I + M)^k$  will be strictly positive. Let  $P = (I + M)^k$  be such a positive matrix; note that (3.1) also implies that matrix  $P = (I + M)^k$  satisfies  $PM = MP$ .

## 程序代与代做 CS 编程辅导

Assume that  $t \leq f(\mathbf{y})$ , i.e., that  $t$  satisfies  $t\mathbf{y} \leq M\mathbf{y}$ ; multiplying both sides by the positive matrix  $P$  we obtain that  $t$  also satisfies

$$tP\mathbf{y} \leq PM\mathbf{y} = MP\mathbf{y};$$

thus,



$$f(P\mathbf{y}) \geq f(\mathbf{y}).$$

Since  $f(\mathbf{y})\mathbf{y} \leq M\mathbf{y}$ , if

if  $f(\mathbf{y})$  is not an eigenvalue of  $M$ , then, since  $P$  is positive,

$$\mathbf{y} < PM\mathbf{y} = MP\mathbf{y}$$

and consequently  $f(P\mathbf{y}) > f(\mathbf{y})$ .

Since every linear operator is continuous and  $S^+$  is compact, the image  $P(S^+)$  is also a compact set as well and  $P(S^+) > 0$ . However,  $\mathcal{J}$  is obviously continuous in a neighbourhood of all points  $\mathbf{y} > 0$ ; thus, it is continuous on  $P(S^+)$ . Consequently, it attains a value  $m$  which is maximal for all  $\mathbf{y} \in P(S^+)$ ,

$$m = \max_{\mathbf{y} \in P(S^+)} f(\mathbf{y}).$$

## Assignment Project Exam Help

Since for every  $\mathbf{y} \geq 0$  we have  $f(\mathbf{y}) \leq f(P\mathbf{y}) \leq m$  we obtain that  $f(\mathbf{y}) \leq m$  for every  $\mathbf{y} \in S^+$  as well; this in turn implies that  $f(\mathbf{y}) \leq m$  for every  $\mathbf{y} \in V^+$ . However, we have shown that whenever  $\mathbf{y}$  is not an eigenvector of  $M$ , then  $f(\mathbf{y}) < f(P\mathbf{y})$ ; thus, we conclude that  $f$  must attain its maximum on  $S^+$  at an eigenvector of  $M$ ; since  $\mathbf{x} \in V^+$ , we have  $\mathbf{x} \neq 0$ .

**Email: tutorcs@163.com**

Since  $f(\mathbf{x})\mathbf{x} = M\mathbf{x}$ ,  $\mathbf{x}$  is also an eigenvector for all  $M^j$  corresponding to the eigenvalue  $(f(\mathbf{x}))^j$  and thus is also an eigenvector of  $P$ :

**QQ: 749389476**

$$P\mathbf{x} = (I + M)^k \mathbf{x} = \sum_{i=0}^k \binom{k}{i} M^i \mathbf{x} = \sum_{i=0}^k \binom{k}{i} (f(\mathbf{x}))^i \mathbf{x} = \left( \sum_{i=0}^k \binom{k}{i} (f(\mathbf{x}))^i \right) \mathbf{x} = (1 + f(\mathbf{x}))^k \mathbf{x}$$

Since  $P$  is positive  $(1 + f(\mathbf{x}))^k > P\mathbf{x} > 0$  which implies that  $\mathbf{x} \neq 0$ .

**https://tutorcs.com**

Let us now consider any other eigenvector  $\mathbf{y}$  corresponding to an eigenvalue  $\lambda$ , i.e., such that  $\lambda\mathbf{y} = M\mathbf{y}$ . Let us denote by  $|\mathbf{y}|$  the vector obtained by taking the absolute values of all coordinates of  $\mathbf{y}$ ; then, since  $M \geq 0$ , we have  $|\lambda||\mathbf{y}| = |\lambda\mathbf{y}| = |M\mathbf{y}| \leq |M||\mathbf{y}| = M|\mathbf{y}|$ . However, this implies that

$$|\lambda| \leq f(|\mathbf{y}|) \leq m = \max_{\mathbf{y} \in V^+} f(\mathbf{y}).$$

Thus,  $m \geq 0$  is an eigenvalue of the largest possible absolute value; note that  $m > 0$  because if  $m = 0$  then all eigenvalues of  $M$  would be 0; thus, the characteristic polynomial of  $M$  would be just  $P(x) = x^n$ . By the Cayley - Hamilton theorem we would have  $M^n = 0$  (i.e.,  $M$  would be nilpotent); thus, for all  $m \geq k$  also  $M^m = 0$ . However, this would contradict irreducibility of  $M$ , because if  $(M^k)_{1,1} > 0$  for some  $k$ , then also  $(M^{k+l})_{1,1} > 0$  for all integers  $l \geq 1$ .

Consider now  $M^\top$ ; then we can apply the previous results to conclude that  $M^\top$  has a positive eigenvector  $\mathbf{q}$  corresponding to the maximal positive eigenvalue. Since  $M$  and  $M^\top$  have the same eigenvalues, we conclude that  $\mathbf{q}$  corresponds to  $\lambda_{\max}$ . But  $M^\top \mathbf{q} = \lambda_{\max} \mathbf{q}$  implies  $\mathbf{q}^\top M = \lambda_{\max} \mathbf{q}^\top$ , i.e.,  $\mathbf{q}^\top$  is a lefthand eigenvector of  $M$  which corresponds to the eigenvalue  $\lambda_{\max}$ .

Assume now that  $\mathbf{y}$  is another eigenvector which satisfies  $\mathbf{y} \geq 0$ , and  $\mu\mathbf{y} = M\mathbf{y}$ . Then

$$\lambda_{\max} \mathbf{q}^\top \mathbf{y} = \mathbf{q}^\top M \mathbf{y} = \mu \mathbf{q}^\top \mathbf{y}$$

Since  $\mathbf{q}^\top > 0$  and  $\mathbf{y} \geq 0$  we obtain  $\mathbf{q}^\top \mathbf{y} > 0$  and thus  $\mu = \lambda_{\max}$ . Moreover,

$$0 < P\mathbf{y} = (1 + \lambda_{\max})^k \mathbf{y}$$

which implies that  $\mathbf{y} > 0$ . Thus, every non negative eigenvector must be strictly positive and must correspond to  $\lambda_{\max}$ .  $\square$

We now want to prove that for a positive matrix  $A$ ,  $\lambda_{max}$  is the only eigenvalue such that  $|\lambda| = \rho(M)$ . For this we first need to prove a few technical lemmas about positive matrices.

**Lemma 3.3** Assume that  $A > 0$  is a positive matrix and  $\mathbf{y} \neq 0$  its eigenvector corresponding to an eigenvalue  $\lambda$  such that  $|\lambda| = \lambda_{max}$ . Then  $T^k \mathbf{y}$  is an eigenvector corresponding to  $\lambda_{max}$ ,



and  $|\mathbf{y}| > 0$ .

**Proof:** We have

$$|\mathbf{y}| = |\lambda \mathbf{y}| = |\lambda_{max} \mathbf{y}|,$$

thus,  $\mathbf{u} = A|\mathbf{y}| - \lambda_{max}|\mathbf{y}|$

$$= |\lambda \mathbf{y}| = |A\mathbf{y}| \leq |A| |\mathbf{y}| = A|\mathbf{y}|;$$

and let us set  $\mathbf{v} = A|\mathbf{y}| > 0$ ; then we obtain

$$0 < A\mathbf{u} = A(\mathbf{v} - \lambda_{max}|\mathbf{y}|) = A\mathbf{v} - \lambda_{max}\mathbf{v}$$

which would imply  $\lambda_{max}\mathbf{v} < A\mathbf{v}$  which would contradict the fact that  $\lambda_{max} = \max_{\mathbf{y} \in V^+} f(\mathbf{y})$ . Thus  $\mathbf{u} = 0$  and consequently  $\lambda_{max}|\mathbf{y}| = A|\mathbf{y}| > 0$  which also implies  $|\mathbf{y}| > 0$ .  $\square$

WeChat: cstutorcs

We now show that  $\mathbf{y}$  from the previous lemma must be just a rotation of  $|\mathbf{y}|$ .

**Lemma 3.4** Assume that  $A > 0$  is a positive matrix and  $\mathbf{y} \neq 0$  its eigenvector corresponding to an eigenvalue  $\lambda$  such that  $|\lambda| = \lambda_{max}$ . Then for some  $\phi \in \mathbb{R}$  we have  $e^{i\phi}\mathbf{y} = |\mathbf{y}| > 0$ .

**Proof:** Since

Email: tutorcs@163.com

and since by the previous lemma we have  $A|\mathbf{y}| = \lambda_{max}|\mathbf{y}|$  and  $|\mathbf{y}| > 0$ , we obtain  $|A\mathbf{y}| = A|\mathbf{y}|$ . Thus, for every  $k$ ,  $1 \leq k \leq n$ , we have  $|\sum_{p=1}^n A_{kp}y_p| = \sum_{p=1}^n A_{kp}|y_p| = \sum_{p=1}^n |A_{kp}y_p|$ . This implies that complex numbers  $A_{kp}y_p$  must all lie on the same ray out of the origin, i.e., the arguments of all complex numbers  $A_{kp}y_p$  must be equal. Since  $A_{kp} > 0$ , this implies that also arguments of all  $y_p$  must be equal. Let  $\phi = -\arg y_1 = -\arg y_2 = \dots = -\arg y_n$ ; then  $e^{i\phi}y_p > 0$  for all  $p$ ,  $1 \leq p \leq n$ , and so we get that  $e^{i\phi}\mathbf{y} > 0$ .  $\square$

QQ: 749389476

We can now show that for a positive matrix  $A$ ,  $\lambda_{max}$  is the only eigenvalue such that  $|\lambda| = \rho(M)$ .

**Lemma 3.5** If  $A > 0$  is a positive matrix, then every eigenvalue  $\lambda \neq \lambda_{max}$  satisfies  $|\lambda| < \lambda_{max}$ .

**Proof:** Since  $|\lambda| \leq \lambda_{max}$ , so it is enough to show that  $|\lambda| = \lambda_{max}$  implies  $\lambda = \lambda_{max}$ . However,  $|\lambda| = \lambda_{max}$  and  $A\mathbf{y} = \lambda\mathbf{y}$  implies that, by the previous lemma, for some  $\phi$  we have  $e^{i\phi}\mathbf{y} > 0$ . Also,  $A\mathbf{y} = \lambda\mathbf{y}$  implies  $Ae^{i\phi}\mathbf{y} = \lambda e^{i\phi}\mathbf{y}$ . But we have proved that all positive eigenvectors must correspond to the eigenvalue  $\lambda = \lambda_{max}$ .  $\square$

Since the eigenvalues of  $A^k$  are  $k^{th}$  powers of the eigenvalues of  $A$ , the above lemma immediately generalises to primitive matrices:

**Theorem 3.6** If  $A \geq 0$  is a primitive matrix, then every eigenvalue  $\lambda \neq \lambda_{max}$  satisfies  $|\lambda| < \lambda_{max}$ .

We now prove that the geometric multiplicity of  $\lambda_{max}$  for positive matrices  $A$  is 1.

**Lemma 3.7** Let  $A > 0$  be a positive matrix, and  $\mathbf{u} \neq 0$  and  $\mathbf{v} \neq 0$  be two eigenvectors for the eigenvalue  $\lambda_{max}$ . Then there exists a constant  $c \in \mathbb{C}$  such that  $\mathbf{u} = c\mathbf{v}$ .

**Proof:** Assume that  $\mathbf{u}$  is not a multiple of  $\mathbf{v}$ . By Lemma 3.4 for some  $\phi_u$  and  $\phi_v$  we have  $\mathbf{p} = e^{i\phi_u}\mathbf{u} > 0$  and  $\mathbf{q} = e^{i\phi_v}\mathbf{v} > 0$ . Let

$$\beta = \min \left\{ \frac{q_i}{p_i} : 1 \leq i \leq n \right\}; \quad \mathbf{r} = \mathbf{q} - \beta \mathbf{p}.$$

Then  $\mathbf{r} \geq 0$  but it is NOT the case that  $\mathbf{r} > 0$ . However, if  $\mathbf{r} \neq 0$  then  $\mathbf{r}$  is also an eigenvector for  $\lambda_{max}$  because it is a linear combination of eigenvectors corresponding to  $\lambda_{max}$ . Thus,  $\mathbf{r} = 1/\lambda_{max}A\mathbf{r} > 0$ , which is a contradiction. Consequently,  $\mathbf{r} = 0$ , and so  $\mathbf{q} = \beta \mathbf{p}$ , i.e.,  $e^{i\phi_v}\mathbf{v} = \beta e^{i\phi_u}\mathbf{u}$ , which implies  $\mathbf{v} = \beta e^{i(\phi_u - \phi_v)}\mathbf{u}$ , contrary to our assumption.  $\square$

Since a vector  $\mathbf{v}$  is an eigenvector of a matrix  $M$  corresponding to an eigenvalue  $\lambda$  just in case  $\mathbf{v}$  is an eigenvector of  $M^k$  corresponding to the eigenvalue  $\lambda^k$ , the above lemma generalises to primitive matrices.

**Theorem 3.8** If  $A \geq 0$  is a primitive matrix, then eigenvalue  $\lambda_{\max} > \rho(A)$  has geometric multiplicity 1.<sup>1</sup>

We now strengthen Theorem 3.2 for the case of primitive matrices.

**Theorem 3.9** 1. Let  $M$  be a primitive matrix; then  $M$  has a real positive eigenvalue  $\lambda_{\max} > 0$  of geometric multiplicity 1.

2. Let  $\mathbf{q}$  and  $\mathbf{x}$  be a left and right eigenvectors of  $M$  corresponding to  $\lambda_{\max}$ , respectively, chosen so that  $\mathbf{x}^T \mathbf{q} = 1$ . Let also  $L = \mathbf{x}\mathbf{q}^T$  be the matrix of rank 1 corresponding to  $\lambda_{\max}$ ; consider matrix  $L = \mathbf{x}\mathbf{q}^T$  of rank 1, and choose  $\mathbf{q}$  and  $\mathbf{x}$  so that  $\mathbf{x}^T \mathbf{q} = 1$ . It is easy to see that  $\mathbf{x}^T \mathbf{q} = 1$  implies that  $L\mathbf{x} = \mathbf{x}$ ,  $\mathbf{q}^T L = \mathbf{q}^T$  and that for all integers  $m \geq 1$ ,  $L^m = L$ . Also,



**Proof:** Note that 1 follows from Theorem 3.2 and 3.8. To prove 2, let again  $\mathbf{q}$  and  $\mathbf{x}$  be the left and the right eigenvector, respectively, corresponding to  $\lambda_{\max}$ ; consider matrix  $L = \mathbf{x}\mathbf{q}^T$  of rank 1, and choose  $\mathbf{q}$  and  $\mathbf{x}$  so that  $\mathbf{x}^T \mathbf{q} = 1$ . It is easy to see that  $\mathbf{x}^T \mathbf{q} = 1$  implies that  $L\mathbf{x} = \mathbf{x}$ ,  $\mathbf{q}^T L = \mathbf{q}^T$  and that for all integers  $m \geq 1$ ,  $L^m = L$ . Also,

WeChat: cstutorcs

$$ML = M\mathbf{x}\mathbf{q}^T = \lambda_{\max}\mathbf{x}\mathbf{q}^T = \lambda_{\max}L$$

$$LM = \mathbf{x}\mathbf{q}^T M = \mathbf{x}\lambda_{\max}\mathbf{q}^T = \lambda_{\max}L.$$

Thus,  $ML = LM$ , which, together with  $L^m = L$ , implies  $ML^m = LM^m = (LM)^m = \lambda_{\max}^m L$ .

We also have  $L(M - \lambda_{\max}L) = LM - \lambda_{\max}L^2 = \lambda_{\max}L - \lambda_{\max}L = 0$ .

Email: tutorcs@163.com

$$(M - \lambda_{\max}L)^m = M^m - \lambda_{\max}^m L. \quad (3.2)$$

QQ: 749389476

We now show that every non-zero eigenvalue  $\mu \neq 0$  of  $M - \lambda_{\max}L$  is also an eigenvalue of  $M$ , and every eigenvector of  $M - \lambda_{\max}L$  corresponding to an eigenvalue  $\mu$  is also an eigenvector of  $M$ . Thus, assume that  $(M - \lambda_{\max}L)\mathbf{z} = \mu\mathbf{z}$ . Multiplying both sides of this equality by  $L$  we obtain that  $\mu L\mathbf{z} = L(M - \lambda_{\max}L)\mathbf{z} = 0$ . Since  $\mu \neq 0$ , we obtain  $L\mathbf{z} = 0$ . This implies that  $\mu\mathbf{z} = (M - \lambda_{\max}L)\mathbf{z} = M\mathbf{z}$ , and consequently  $\mu$  is also an eigenvalue of  $M$  and  $\mathbf{z}$  is an eigenvector of  $M$ .

We now show that  $\lambda_{\max}$  cannot be an eigenvalue of  $M - \lambda_{\max}L$ . To see this, assume the opposite, i.e., that  $\lambda_{\max}$  is an eigenvalue of  $M - \lambda_{\max}L$  and that  $\mathbf{z}$  is the corresponding eigenvector. But we have just shown that in this case  $\mathbf{z}$  would also be an eigenvector of  $M$  which corresponds to  $\lambda_{\max}$ . Since we have shown that the geometric multiplicity of  $\lambda_{\max} > 0$  is one,  $\mathbf{z} = \alpha\mathbf{x}$ , for some  $\alpha \neq 0$ . This implies that  $(M - \lambda_{\max}L)\alpha\mathbf{x} = \lambda_{\max}\alpha\mathbf{x}$ , so also  $(M - \lambda_{\max}L)\mathbf{x} = \lambda_{\max}\mathbf{x}$  which implies  $\lambda_{\max}\mathbf{x} = M\mathbf{x} - \lambda_{\max}L\mathbf{x} = \lambda_{\max}\mathbf{x} - \lambda_{\max}\mathbf{x} = 0$  which is a contradiction.

To summarise, we have just shown that the eigenvalue  $\lambda^*$  of  $M - \lambda_{\max}L$  of the largest absolute value is at most by absolute value the second largest eigenvalue of  $M$ .

From (3.2) we have

$$\left(\frac{M}{\lambda_{\max}} - L\right)^m = \left(\frac{M}{\lambda_{\max}}\right)^m - L, \quad (3.3)$$

and the absolute value of any eigenvalue of  $M/\lambda_{\max} - L$  is at most  $|\lambda^*|/|\lambda_{\max}| < 1$ .

One can show that for any matrix  $A$  if all of its eigenvalues  $\lambda$  satisfy  $|\lambda| < r < 1$ , then  $A^m \rightarrow 0$ , and for some constant  $C > 0$ ,  $|(A^m)_{ij}| < Cr^m$ . (This is true in general, but it is particularly easy to see for matrices which have  $n$  linearly independent eigenvectors. As we know from linear algebra, if  $A$  has  $n$  linearly independent eigenvectors, then  $A = Q\Lambda Q^{-1}$ , where  $\Lambda$  is a diagonal matrix with the eigenvalues on the diagonal and zeros elsewhere, and  $Q$  is a matrix whose columns are the right eigenvectors of  $A$ . This implies  $A^m = Q\Lambda^m Q^{-1}$ ; since  $\Lambda^k$  has  $\lambda_i^m$  on the diagonal and  $|\lambda_i|^m < r^m$ , the above bound easily follows.)

Applying this to matrix  $M/\lambda_{\max} - L$  completes our proof of 2 of the Theorem.  $\square$

<sup>1</sup>One can show that algebraic multiplicity of  $\lambda_{\max}$  is also 1, but we do not need this.

## 4 Markov Chains revisited

We now want to apply the above machinery to show that every aperiodic and irreducible Markov Chain has a unique stationary probability distribution.

**Theorem 4.1** *The transition matrix  $M$  of an irreducible aperiodic Markov chain is primitive.*

Let  $P$  be any state; since  $P$  is a vertex in the underlying graph there exists a directed loop  $\pi_0$  of some length  $l_0$  which contains vertex  $P$  at least once. Let  $p_1, \dots, p_k$  be the prime divisors of  $l_0$ . Since  $M$  is aperiodic, there exist loops  $\pi_1, \dots, \pi_k$  containing  $P$  at least once such that  $l_i$  is not divisible by  $p_i$ . Thus,  $\gcd(l_0, l_1, \dots, l_k) = 1$ . It is a theorem in number theory that for every integer  $n > K_P$  there exists an integer  $K_P > 0$  such that every integer  $n = a_0l_0 + a_1l_1 + \dots + a_kl_k$  with all of the coefficients  $a_0, \dots, a_k$  non negative integers. The smallest such number  $K_P$  is called the *Frobenius number* for the sequence  $l_0, \dots, l_k$ . We now prove that for every  $n > K_P$  there is a loop containing  $P$  of length precisely  $n$ . The loop is constructed as follows: first we go around the loop  $\pi_0$  of length  $l_0$  exactly  $a_0$  many times, then we go through the loop  $\pi_1$  of length  $l_1$  exactly  $a_1$  many times, etc.; after going through the loop  $\pi_k$  exactly  $a_k$  many times, this will in total constitute a loop through  $P$  of length exactly  $a_0l_0 + a_1l_1 + \dots + a_kl_k = n$ . Let  $m$  be the total number of states in the chain; we now take  $N = \max\{K_P : P \text{ is a state of the Markov Chain}\} + m$ . Then for every two states  $P$  and  $Q$ , by irreducibility, there exists a (non self intersecting) path from  $P$  to  $Q$  of length  $L \leq m - 1$ ; after reaching  $Q$  we go around a cycle containing  $Q$  of length exactly  $N - L$  which exists by the definition of  $N$ ; since the total length of the path is  $N$ , and since a path of length  $N$  exists just in case  $(M^N)_{P,Q} > 0$ , given that  $P$  and  $Q$  were arbitrary, we conclude that  $M^N > 0$ .



underlying graph there exists a directed loop  $\pi_0$  of some length  $l_0$  which contains vertex  $P$  at least once. Let  $p_1, \dots, p_k$  be the prime divisors of  $l_0$ . Since  $M$  is aperiodic, there exist loops  $\pi_1, \dots, \pi_k$  containing  $P$  at least once such that  $l_i$  is not divisible by  $p_i$ . Thus,  $\gcd(l_0, l_1, \dots, l_k) = 1$ . It is a theorem in number theory that for every integer  $n > K_P$  there exists an integer  $K_P > 0$  such that every integer  $n = a_0l_0 + a_1l_1 + \dots + a_kl_k$  with all of the coefficients  $a_0, \dots, a_k$  non negative integers. The smallest such number  $K_P$  is called the *Frobenius number* for the sequence  $l_0, \dots, l_k$ . We now prove that for every  $n > K_P$  there is a loop containing  $P$  of length precisely  $n$ . The loop is constructed as follows: first we go around the loop  $\pi_0$  of length  $l_0$  exactly  $a_0$  many times, then we go through the loop  $\pi_1$  of length  $l_1$  exactly  $a_1$  many times, etc.; after going through the loop  $\pi_k$  exactly  $a_k$  many times, this will in total constitute a loop through  $P$  of length exactly  $a_0l_0 + a_1l_1 + \dots + a_kl_k = n$ . Let  $m$  be the total number of states in the chain; we now take  $N = \max\{K_P : P \text{ is a state of the Markov Chain}\} + m$ . Then for every two states  $P$  and  $Q$ , by irreducibility, there exists a (non self intersecting) path from  $P$  to  $Q$  of length  $L \leq m - 1$ ; after reaching  $Q$  we go around a cycle containing  $Q$  of length exactly  $N - L$  which exists by the definition of  $N$ ; since the total length of the path is  $N$ , and since a path of length  $N$  exists just in case  $(M^N)_{P,Q} > 0$ , given that  $P$  and  $Q$  were arbitrary, we conclude that  $M^N > 0$ .

**WeChat: cstutorcs Assignment Project Exam Help**

**Homework Problem 3.** Prove that if in an irreducible Markov Chain one state is aperiodic, then all states must be aperiodic.

**Email: tutorcs@163.com**

We can now prove the main theorem about the Markov chains, Theorem 2.1:

**Theorem 2.1** *Any finite, irreducible and aperiodic Markov chain has the following properties:*

1. *Let the matrix  $M$  correspond to such a Markov process; for every initial probability distribution of states  $\mathbf{q}^{(0)}$  the value of  $(\mathbf{q}^{(t)})^\top = (\mathbf{q}^{(0)})^\top M^t$  converges as  $t \rightarrow \infty$  to a unique stationary distribution  $\mathbf{q}$  which is independent of  $\mathbf{q}_0$ , i.e., converges to the unique distribution  $\mathbf{q}$  which satisfies  $\mathbf{q}^\top = \mathbf{q}^\top G$ .*
2. *Let  $N(P_i, T)$  be the number of times the system has been in the state  $P_i$  during  $T$  many transitions of such a Markov chain; then*

$$\lim_{T \rightarrow \infty} \frac{N(P_i, T)}{T} = q_i.$$

**Proof:** By the previous theorem matrix  $M$  of a finite, irreducible and aperiodic Markov chain is primitive. Since  $M$  is also stochastic, it is easy to see that for  $\mathbf{x}^\top = (1, 1, \dots, 1)$  we have  $M\mathbf{x} = \mathbf{x}$ . Thus, 1 is an eigenvalue of  $M$  with  $\mathbf{x} > 0$  as a corresponding positive eigenvector. As we have shown, every non negative eigenvector corresponds to  $\lambda_{\max}$ . From linear algebra we know that the eigenvectors corresponding to distinct eigenvalues are linearly independent, so we can conclude that  $\lambda_{\max}$  must be equal to 1. Let  $\mathbf{q} > 0$  be the left eigenvector corresponding to the eigenvalue  $\lambda_{\max} = 1$  chosen so that  $\mathbf{x}^\top \mathbf{q} = 1$ . Then  $\mathbf{q}^\top M = \mathbf{q}^\top$ ; thus,  $\mathbf{q}$  is the stationary probability distribution for  $M$  as required by the theorem.

Moreover, as we have shown in Theorem 3.9,  $\lim_{t \rightarrow \infty} M^t = L = \mathbf{x}\mathbf{q}^\top$ . Given what  $\mathbf{x}$  is, matrix  $L$  has the following form

$$L = \begin{pmatrix} q_1 & q_2 & \dots & \dots & q_{n-1} & q_n \\ q_1 & q_2 & \dots & \dots & q_{n-1} & q_n \\ \vdots & \vdots & & & \vdots & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ q_1 & q_2 & \dots & \dots & q_{n-1} & q_n \\ q_1 & q_2 & \dots & \dots & q_{n-1} & q_n \end{pmatrix}$$

Thus, for every positive vector  $\mathbf{q}_0$  whose entries sum up to 1 we obtain

$$\lim_{t \rightarrow \infty} \mathbf{q}_0^\top M^t = \mathbf{q}_0^\top L = \mathbf{q}_0^\top \mathbf{x}\mathbf{q}^\top = \mathbf{q}^\top$$

Note that Theorem 3.9 implies

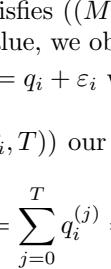
$$(M - L)^m = M^m - L = M^m - \mathbf{x}\mathbf{q}^\top$$

Let us denote  $\mathbf{q}_0^\top M^m$  by  $(\mathbf{q}^{(m)})^\top$ , since

$$\mathbf{q}_0^\top (M - L)^m = \mathbf{q}_0^\top (M^m - \mathbf{x}\mathbf{q}^\top) = \mathbf{q}_0^\top M^m - \mathbf{q}_0^\top \mathbf{x}\mathbf{q}^\top = (\mathbf{q}^{(m)})^\top - \mathbf{q}^\top$$

and since every entry of  $\mathbf{t}$  satisfies  $((M - L)^m)_{ij} < cr^m$  for any  $|\lambda^*| < r < 1$ , where  $\lambda^*$  is the second largest eigenvalue of  $M$ , we obtain that for some  $C$  and for every  $i$ ,  $1 \leq i \leq n$  we have  $|(\mathbf{q}^{(m)})_i - q_i| < Cr^m$ .

Clearly, the expected time from 1 to  $T$  is

   $P_i, T)$ ) our Markov chain will be in state  $P_i$  over a period of

$$(\mathbf{q}^{(m)})_i = \sum_{j=0}^T q_i^{(j)} = Tq_i + \sum_{j=0}^T \varepsilon_i$$

Thus we obtain

$$\left| E\left(\frac{N(P_i, T)}{T}\right) - q_i \right| = \left| \frac{C \sum_{j=0}^T \varepsilon_j}{T} \right| \leq \frac{C \sum_{j=0}^T |\varepsilon_j|}{T} \leq \frac{C \sum_{j=0}^T r^j}{T} < \frac{C}{T(1-r)} \rightarrow 0.$$

Since the expected value  $E\left(\frac{N(P_i, T)}{T}\right)$  converges to  $q_i$ , and since by the strong law of large numbers the value of  $\frac{N(P_i, T)}{T}$  converges almost surely to its expected value  $E\left(\frac{N(P_i, T)}{T}\right)$ , we get that  $\frac{N(P_i, T)}{T}$  must also converge to  $q_i$ .  $\square$

## Email: tutorcs@163.com

**Homework Problem 4.** You are watching traffic on a busy road and you notice that on average three out of every four trucks on the road are followed by a car, while only one out of every five cars is followed by a truck. What fraction of vehicles on the road are trucks?

 QQ: 749389476

*Hint: you might want to use Theorem 2.1.*

## 5 Google PageRank revisited

Recall that the Google matrix is of the form

$$G = \alpha G_2 + \frac{1-\alpha}{M} \mathbf{e}\mathbf{e}^\top, \quad \text{with } G_2 = G_1 + \frac{1}{M} \mathbf{d}\mathbf{e}^\top; \quad (5.1)$$

Note that both  $G_2$  and  $G$  are stochastic matrices. More over,  $G$  is not only a primitive matrix, but it is in fact a positive matrix. Thus, the Markov Chain theorem applies and for some positive vector  $\mathbf{q}$  we have  $\mathbf{q}G = \mathbf{q}$ , and our random surfer will visit a site  $P_i$  on average  $q_i T$  many times during a surfing session with  $T$  many clicks, regardless of the particular surfing history.

We now want to estimate the convergence rate of our iterative procedure for obtaining the PageRank stationary distribution  $\mathbf{q}$ . We first prove the following lemma.

**Lemma 5.1** *Every eigenvalue  $\mu$  of a non negative stochastic matrix  $M$  satisfies  $|\mu| \leq 1$ .*

**Proof:** Let  $\mathbf{y} \neq 0$  be an eigenvector corresponding to an eigenvalue  $\mu$  of an  $n \times n$  stochastic matrix  $M$ . Let also  $k$  be such that  $|y_j| \leq |y_k|$  for all  $1 \leq j \leq n$ . Then looking at the  $k^{\text{th}}$  coordinate of both sides of  $\mu\mathbf{y} = M\mathbf{y}$  we obtain

$$|\mu| |y_k| = |\mu y_k| = \left| \sum_{j=0}^n M_{kj} y_j \right| \leq \sum_{j=0}^n M_{kj} |y_j| \leq \sum_{j=0}^n M_{kj} |y_k| = |y_k|$$

and thus  $|\mu| \leq 1$ .  $\square$

Clearly, for  $\mathbf{x}^\top = (1, 1, \dots, 1)$ , since  $G_2$  is row stochastic non negative matrix we have  $G_2 \mathbf{x} = \mathbf{x}$ . Let  $Q$  be any invertible matrix whose first column is  $\mathbf{x}$ , so  $Q = (\mathbf{x} \ X)$  for some  $n \times (n-1)$  matrix  $X$ , and let us write  $Q^{-1}$  as:

$$Q^{-1} = \begin{pmatrix} \mathbf{y}^\top \\ Y \end{pmatrix}$$

where  $\mathbf{y}$  is a vector and  $Y$  is an  $(n - 1) \times n$  matrix. Note that, since  $Q^{-1}Q = I_n$  (the identity matrix of size  $n \times n$ ),

# 程序代写代做CS编程辅导

$$Q^{-1}Q = \begin{pmatrix} \mathbf{y}^\top \\ Y \end{pmatrix} (\mathbf{x}^\top X) = \begin{pmatrix} \mathbf{y}^\top \mathbf{x} & \mathbf{y}^\top X \\ Y \mathbf{x} & YX \end{pmatrix} = I_n.$$

Thus,

$$\mathbf{y}^\top \mathbf{x} = 1; \quad \text{QR code: } \begin{array}{|c|} \hline \text{Tutor CS} \\ \hline \end{array} \quad Y \mathbf{x} = \mathbf{0}; \quad YX = I_{n-1}. \quad (5.2)$$

We also obtain by block matrices:

$$G_2(\mathbf{x}^\top X) = \begin{pmatrix} \mathbf{y}^\top G_2 \mathbf{x} & \mathbf{y}^\top G_2 X \\ Y G_2 \mathbf{x} & Y G_2 X \end{pmatrix}$$

Since  $G_2$  is stochastic,  $G_2 \mathbf{x}^\top X = \mathbf{x}^\top X$

$$Q^{-1}G_2Q = \begin{pmatrix} \mathbf{y}^\top \mathbf{x} & \mathbf{y}^\top G_2 X \\ Y \mathbf{x} & Y G_2 X \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{y}^\top G_2 X \\ \mathbf{0} & Y G_2 X \end{pmatrix} \quad (5.3)$$

## WeChat: cstutorcs

Similarly, for the rank 1 matrix  $\mathbf{x}\mathbf{x}^\top$ , we have

$$Q^{-1}\mathbf{x}\mathbf{x}^\top Q = (Q^{-1}\mathbf{x})(\mathbf{x}^\top Q) = \begin{pmatrix} \mathbf{y}^\top \mathbf{x} \\ Y \mathbf{x} \end{pmatrix} (\mathbf{x}^\top \mathbf{x} \quad \mathbf{x}^\top X) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1 \quad \mathbf{x}^\top X) = \begin{pmatrix} 1 & \mathbf{x}^\top X \\ \mathbf{0} & 0 \end{pmatrix} \quad (5.4)$$

Combining (5.3) and (5.4), we obtain

$$Q^{-1}GQ = Q^{-1} \left( \alpha G_2 + \frac{1-\alpha}{M} \mathbf{x}\mathbf{x}^\top \right) Q = \begin{pmatrix} 1 & \alpha \mathbf{y}^\top G_2 X + (1-\alpha)\mathbf{x}^\top X \\ \mathbf{0} & \alpha Y G_2 X \end{pmatrix} \quad (5.5)$$

## Assignment Project Exam Help

Note that matrices  $G$  and  $Q^{-1}GQ$  share the same characteristic polynomial:

$$\text{Det}(G - \lambda I) = \text{Det}(Q^{-1}) \text{Det}(G - \lambda I) \text{Det}(Q) = \text{Det}(Q^{-1}(G - \lambda I)Q) = \text{Det}(Q^{-1}GQ - \lambda I).$$

Thus, the eigenvalues of  $G$  are the same as eigenvalues of  $Q^{-1}GQ$  and the eigenvalues of  $G_2$  are the same as eigenvalues of  $Q^{-1}G_2Q$ . Let  $1, \lambda_2, \lambda_3, \dots, \lambda_n$  be all the eigenvalues of  $G_2$ . We now use the fact that the set of eigenvalues (including the multiplicities) of a matrix of the form

## https://tutorcs.com

is the union of the set of eigenvalues of the matrix  $A_1$  and the set of eigenvalues of  $A_2$  to conclude that (5.3) implies that the eigenvalues of  $YG_2X$  are  $\lambda_2, \lambda_3, \dots, \lambda_n$ . (5.5) now implies that the eigenvalues of matrix  $G$  are  $1, \alpha\lambda_2, \alpha\lambda_3, \dots, \alpha\lambda_n$ . Since  $G_2$  is stochastic, all of its eigenvalues satisfy  $|\lambda_i| \leq 1$ . Thus, the second largest (by absolute value) eigenvalue of  $G$  is  $\alpha\lambda_2$  which satisfies  $|\alpha\lambda_2| = |\alpha\lambda_2| \leq \alpha$ .

Consequently, by Theorem 3.9 the rate of convergence of the power method for evaluating the PageRank is  $O(r^n)$  for any  $r$  such that  $\alpha < r$ , i.e., (essentially)  $O(\alpha^n)$ . For  $\alpha = 0.85$  we have  $0.85^{50} \approx 0.000296$  which appears to be sufficient to determine the relative ranking of webpages, even if the values are not quite exact.

As we have mentioned, larger values of  $\alpha$  make not only the convergence of the power method slower, but also such obtained PageRank more sensitive to changes in the structure of the web. Excessive sensitivity is not a good thing even if the resulting PageRank better reflects the structure of the web at that moment, because if PageRank changed significantly quite often, it would be of dubious value as an indicator of importance of webpages. Thus, striking the right balance in choosing the value of  $\alpha$  is of utmost importance.

### Notes:

1. The PageRank algorithm has also been adapted to other uses. You can find a UNSW homegrown application in the field of computer security in the last reference below, available at the course website. It grew out of two student's major projects for this very class, of Mohsen Rezvani's who came up with the idea and Verica Sekulic's who gave a lovely proof of convergence of the algorithm. So take the major project for this class as a challenge to produce a high quality publishable piece of work!
2. Finding a vector  $\mathbf{q}$  such that  $\mathbf{q}^\top = \mathbf{q}^\top G$  amounts to finding a **fixed point** of the linear mapping which corresponds to  $G$ , i.e., the linear mapping  $\mathbf{v} \rightarrow \mathbf{v}^\top G$ . As we will see, several other algorithms which we will present later are also fixed points algorithms, albeit of non linear mappings, including a voting algorithm which we present next.

## Further Reading

1. Mung Chiang: Networked Life, Cambridge University Press, 2012;
2. Amy Langville and Carl Meyer: Google's PageRank and Beyond, Princeton University Press, 2006;

3. Amy Langville and Carl Meyer: Deeper Inside PageRank,  
<http://meyer.math.fsu.edu/Files/DeeperInsidePR.pdf>;

4. Roger Horn and Charles Johnson: Matrix Analysis;

5. Lecture slides on PageRank:  
 by Shlomo Sternberg,  
<http://isites.harvard.edu/fs/tutorcs/Topic540499.files/1180912pf.pdf>;

6. Mohsen Rezvani, Vahid Khodabakhshian, Saeed Javidi, Naser Ignjatovic, Elisa Bertino, Sanjay Jha, "Interdependent Security Risk Analysis", in: IEEE Transactions on Information Forensics and Security IEEE Transactions on, vol. 10, no. 11, pp. 2325-2339, 2015

The last item which presents a PageRank style algorithm for computer security is available at the class website.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>