# VIDYASAGAR UNIVERSITY



PROJECT SUBMITTED FOR PARTIAL FULFILMENT OF BACHELOR’S DEGREE IN SCIENCE IN STATISTICS HONOURS

* STUDY THE IMPORTANCE OF FOUR BUS STOPS USING THE GOOGLE PAGE RANKIN METHOD

Registration No. 1160533

Session : 2019 – 2020 DEPARTMENT OF STATISTICS HALDIA GOVERNMENT COLLEGE

ACKNOWLEDGEMENT

It gives me tremendous pleasure in acknowledging that valuable assistance extended to me by various people in the successful completion of this program.

I would like to thank all the people who contributed in some way in the course of my project work in the Department of Statistics, Haldia Government College. Firstly, I offer my sincere gratitude to project guide Mr. Tuhinsubhra Bhattacharya , Dr. Shyamsundar Sahoo, Mr. Sibsankar Karan, Mr. Tanmay Maity and Mr. Bijitesh Haldar for continuous support for my project, for their patience, motivation and immense knowledge. His guidance helped me in all the time of computing and writing this project. I could not have imagined a better project guide and mentor for my project work. Last but not least I would like to thank all of my friends who have helped and gave me timely support due to which the project has been completed successfully.

### Subhadip Ghosh

**CONTENTS**

1. INTRODUCTION
2. OBJECTIVES
3. GOOGLE PAGE RANKING METHOD
4. MARKOV MATRIX
5. ASSUMPTIONS
6. DATA COLLECTION
7. METHODOLOGY
8. ANALYSIS & RESULTS
9. CONCLUSION
10. FURTHER STUDIES
11. REFERENCES
12. APPENDIX

# INTRODUCTION:-

## Google page rank :

Page Rank (PR) is an algorithm used by Google Search to rank websites in their search engine results. Page Rank was named after Larry Page, one of the founders of Google. Page Rank is a way of measuring the importance of website pages.According to Google: Page Rank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.This is the application of eigen vector & eigen value .We can order the bus stops ( Nimtouri , Nandakumar , Brajalalchak, Ranichak) by the method of "GOOGLE PAGE RANKING "

# Objectives

* WE HAVE TAKEN FOUR BUS STOPS NAMELY NIMTOURI, BRAJALALCHAK, NANDAKUMAR, RANICAHK.
* WHICH BUS STOP IS MORE IMPORTANT OF OTHERS BUS STOPS .

## 3) Google page ranking method

### Let find a example.

Firstly I took four(4) web pages & names assign 1, 2, 3, 4 . May be pages are Facebook.com, yahoo.co.in,amazon.in, Hotstar . in .

Let, it is possible to go from link number 1 to link number 2,3,4 . same way going to link number 1 & 4 from link number

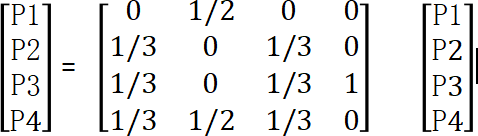
2. and going to link number 2,3,4 from 3 & I can only go from 4 number link to 3 . from the figure 1 , can we say that which web page is most important . Certainly we say that , 3&4 are most important web page according to maximum pages go to this web pages , but it’s wrong because it’s not a ranking .That is why we use Google page ranking method. I want to do that,if we doing browse randomly,then how could we will come t others pages.In the case of number one this rate is P.Alike in the case of number 2 it will be P2.. And in the case of 3 it will be P3. in the case of 4 it will be P4 .

### we can say that

P1 = 0 \* P1 + 1/2 P2 +0 \* P3 + 0 \* P4 P2 = 1/3 P1 + 0 \* P2 + 1/3 P3+ 0 \* P4

### P3 = 1/3 P1 + 0 \* P2 + 1/3 P3+ 1 P4 P4 = 1/3 P1 +1/2 P2 + 1/3 P3 + 0 \* P4

SO, we can write .



So , P =A P

here, P is a eigen vector of the matrix A , where eigen value 1 .

Now ,

From the eigen vertor we want to which web page is most important of four web pages .

## Markov Matrix

In mathematics, a stochastic matrix is a square matrix used to describe the transitions of a Markov chain. Each of its entries is a non negative real number representing a probability.   It is also called a probability matrix, transition

matrix, substitution matrix, or Markov matrix.  The stochastic matrix was first developed by Andrey Markov at the beginning of the 20th century, and has found use throughout a wide variety of scientific fields, including probability theory, statistics, mathematical finance and linear algebra, as well as computer science and population genetics.  There are several different definitions and types of stochastic matrices.

A right stochastic matrix is a real square matrix, with each row summing to 1.

A left stochastic matrix is a real square matrix, with each column summing to 1.

A doubly stochastic matrix is a square matrix of non negative real numbers with each row and column summing to 1.

In the same vein, one may define a stochastic vector (also called probability vector) as a vector whose elements are non negative real numbers which sum to 1. Thus, each row of a right stochastic matrix (or column of a left stochastic matrix) is a stochastic vector. A common convention in English language mathematics literature is to use row vectors of probabilities and right stochastic matrices rather than column vectors of probabilities and left stochastic matrices.

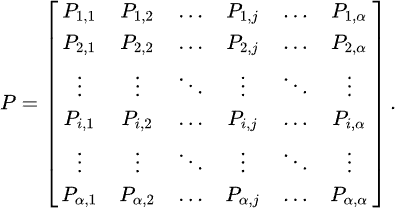
**Properties :**

A stochastic matrix describes a Markov chain Xt

over a finite state space S with cardinality α.

If the probability of moving from i to j in one time step is Pr(j|i)

= Pi,j, the stochastic matrix P is given by using Pi,j as the i-th row and j-th column element, e.g.,



Since the total of transition probability from a state i to all other states must be 1,



thus this matrix is a right stochastic matrix.

The above element wise sum across each row i of P may be more concisely written as P1 = 1, where 1 is the α -dimensional vector of all ones. Using this, it can be seen that the product of two right stochastic matrices P′ and P′′ is also right stochastic: P′ P′′ 1 = P′ (P′′ 1) = P′ 1 = 1. In general, the k th power Pk of a right stochastic matrix P is also right stochastic.

In general, the probability transition of going from any state to another state in a finite Markov chain given by the matrix P in k steps is given by Pk.An initial probability distribution of states, specifying where the system might be initially and with what probabilities, is given as a row vector.A stationary probability vector π is defined as a distribution, written as a row vector, that does not change under application of the transition matrix; that is, it is defined as a probability distribution on the set {1, …, n} which is also a row eigenvector of the probability matrix, associated with eigenvalue 1.



# ASSUMPTION:

p1= Rate of bus riders from other bus stop to Nimtouri

p2= Rate of bus riders from other bus stop to Nandakumar . p3 = Rate of bus riders from other bus stop to Brajalalchak p4 = Rate of bus riders from other bus stop to Ranichak.

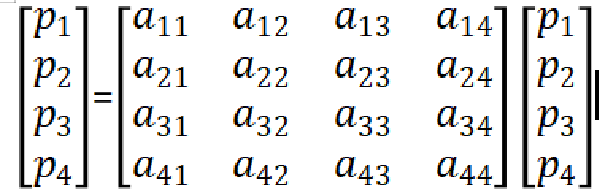
we assing 1 ,2 ,3,4 to the bus stops respectively Nimtouri , Nandakumar , Brajalalchak,

Ranichak.

ai1 = Rate of bus riders from bus stop i to Nimtouri . i =1,2,3,4 ai2=Rate of bus riders from bus stop i to Nandakumar .i=1,2,3,4 ai3=Rate of bus riders from bus stop i to Brajalalchak. i = 1,2,3,4 ai4 = Rate of bus riders from bus stop i to Ranichak. i = 1,2,3,4

We can write the linear equations.

p1 = a11 p1 + a12 p2 +a13 p3 + a14 p4 p2 = a21 p1 + a22 p2 +a23 p3 + a24 p4 p3 = a31 p1 + a32 p2 +a33 p3 + a34 p4 p4 = a41 p1 + a42 p2 +a43 p3 + a44 p4

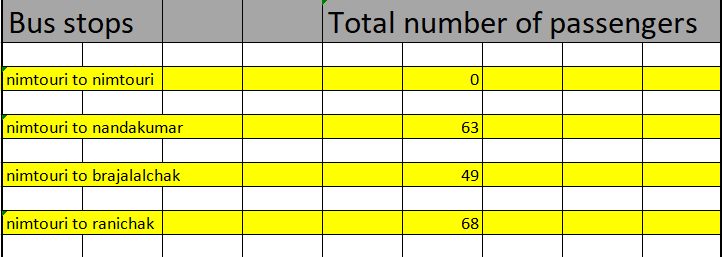


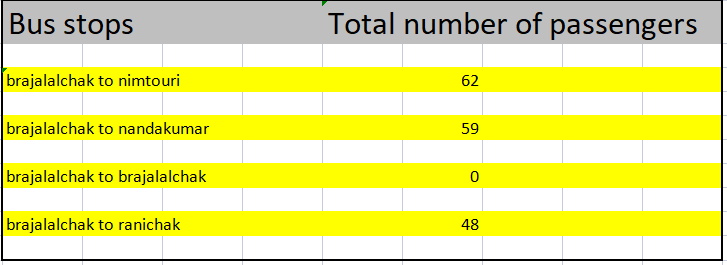
So , 𝑝 =A 𝑝

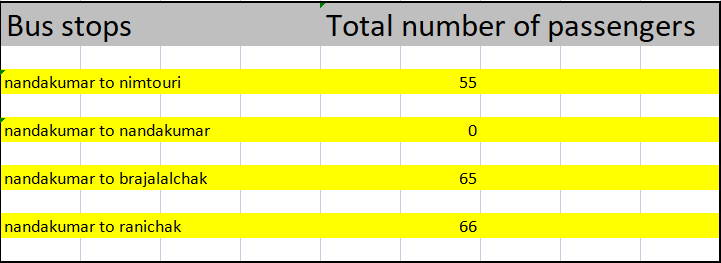
here, 𝑝 is a eigen vector of the matrix A , where eigen value 1 .

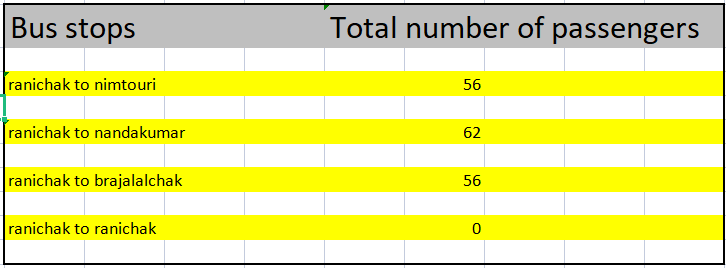
# Data collection :

I have collected data from various students of my college.The data is collected from 100 students..









# METHODOLOGY: -

Total number of passengers complete his/her ride from NIMTOURI (A) = 181

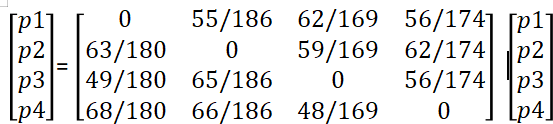
Total number of passengers complete his/her ride from NANDAKUMAR( B) = 186

Total number of passengers complete his/her ride from BRAJALALCHAK(C) = 169

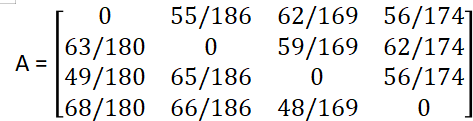
Total number of passengers complete his/her ride from RANICHAK(D) = 174

p1 = 0\* p1 + 55/186 p2 + 62/169 p3 + 56/174 p4

p2 = 63/180 p1 + 0\* p2 + 59/169 p3 + 62/174 p4 p3 = 49/180 p1 + 65/186 p2 + 0\* p3 + 56/174 p4 p4 = 68/180 p1 + 66/ 186 p2 + 48/169 p3 + 0\* p4



Where, A matrix is

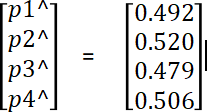


1. Analysis & Results :

Here, p =A p

p is a eigen vector of the matrix A , where eigen value 1 . HERE ,

P =



# CONCLUSION : -

From the eigen vector we found that the most important bus stop is **NANDAKUMAR** of other bus stops . And the less important bus stop is **BRAJALALCHAK** of other bus stops .

# Further Studies :

* We can estimate the importance of rail stations & airports using this method.

# References :

* Linear Algebra : Second Edition By Arnab Chakraborty
* Linear Algebra and Matrix Analysis for Statistics :Sudipto Banerjee & Anindya Roy
* Linear Algebra: Step by Step By Kuldip Singh

# Appendix :

> a1<-c(0,55/186,62/169,56/174)

> a2<-c(63/180,0,59/169,62/174)

> a3<-c(49/180,65/186,0,56/174)

> a4<-c(68/180,66/186,48/169,0)

> a<-rbind(a1,a2,a3,a4)

> eigen(a)