Modified Pagerank Algorithm Based Real-Time Metropolitan Vehicular Traffic Routing Using GPS Crowdsourcing Data

Adithya Guru Vaishnav.S

Abstract: This paper aims at providing a theoretical framework to find an optimized route from any source to destination considering the real-time traffic congestion issues. The distance of various possible routes from the source and destination are calculated and a PathRank is allocated in the descending order of distance to each possible path. Each intermediate locations are considered as nodes of a graph and the edges are represented by real-time traffic flow monitored using GoogleMaps GPS crowdsourcing data. The Page Rank is calculated for each intermediate node. From the values of PageRank and PathRank, the minimum sum term is used to find an optimized route with minimal trade-off between shortest path and real-time traffic.

Index Terms: PageRank Algorithm, Global Positioning System (GPS), Google Maps, Google Crowdsourcing, Nascent Rank algorithm

1. INTRODUCTION

Nowadays, due to the advent of plethora of automobiles in the metropolitan cities, navigation has become cumbersome for the commuters. This is further worsened during peak hours. People tend to get confused to choose an optimized route from the available multiple paths which would guide them to the destination in an optimized time. GoogleMaps has been devised to provide real-time traffic data to the users, beforehand. The purpose of this paper is to further enhance the GoogleMaps services by proposing a nascent algorithm by combining the Google's PageRank Algorithm and the distance measured by Global Positioning System(GPS).The algorithm proposed here has not been practically implemented but the results are simulated using directed graphs analogy on the maps and thereby performing PageRank calculation methodology on them.

2. PAGERANK ALGORITHM

As this paper utilizes the modified form of PageRank Algorithm, a brief description of this algorithm is given. It was proposed by Google founders Sergey Brin and Lawrence Page to optimize the Google search engine by displaying the results based on the order of relevance to the search. This algorithm overcame the text based ranking systems by providing results on the basis of number of hypertext backlinks linked with the particular page. The basic assumption is that the more important websites are bound to get more links from other websites.

- *Adithya Guru Vaishnav. S is currently pursuing bachelor degree in electronics and communication engineering in Kumaraguru College of Technology, Anna University, India. PH-919944722609.*
- *E-mail: adithyagvaishnav@gmail.com*

A webgraph is created assuming all webpages as nodes and the hyperlinks that are attached to them as edges. Each page is assigned a PageRank and it is calculated recursively till the rank converges and based on the converged values the final ranks are allocated. A term called "Damping Factor" has been introduced in order to overcome the dangling effects and unrelated nodes misinterpretations which would result in false PageRanks.

Let us consider a simple webgraph network with three nodes A,B and C.

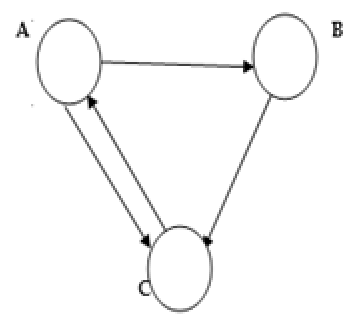

Fig 1. *A webgraph example with 3 nodes A,B ,C.*

The rank of node A can be calculated using the formula,
PR(A)=(1-d)/N+d(PR(T1)/C(T1)+…+PR(Tn)/C(Tn))

Here the term d is the damping factor which is used to overcome the dangling effects[1].

N is the total number of pages on the web.In this case, N=3.

By iterative computational convergence using the above formula, it can be found that,
PR(A)=1.07692308
PR(B)=0.76923077
PR(C)=1.15384615
So, A,B,C are assigned PR2,PR1,PR3 respectively.

In this paper, as we are dealing with real-time places represented as nodes and traffic flow is represented as edges, we use the term Locational Page Rank(LPR) to set the rank for the intermediate locations connected in the map. Moreover, during the calculation of LPR, we use Markov chain transition matrix as dangling effects are negligible in the road map where we are going to perform LPR calculations.

3. GLOBAL POSITIONING SYSTEM

GPS is a space based navigation system which uses satellites to provide location and time information in all weather conditions. It is operated by the Department of Defence of the U.S. Government. This technology relies on satellite data for accurate determination of any location. It consists of satellites, control and monitor stations and receivers. Approximately, three satellites are needed to estimate the 2D position of a particular location on the map. For accurate calculations four or more satellites are required[2]. There are three segments in GPS, Space segment: This consists of 24 satellites oribiting the earth at an altitude of 11,000 nautical miles. Control segment: This consists of control and monitor rooms installed at various places in the U.S. and they help in the accurate determination of the locations. Any discrepancies between the actual orbits and the predicted orbits are transmitted to the satellites. The satellites can broadcast these corrections along with timing data and position so that the GPS receiver can precisely establish the location of each of the satellite it is tracking. User segment: GPS is used for both military purposes as well as civilian uses such as surveying, mapping, timing and navigational purposes.

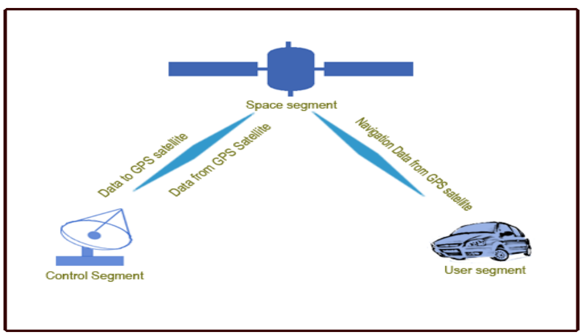

Fig 2. *An overview of GPS segments.*

Using this technology, the distance between each successive node sets are calculated. For instance let the co-ordinates of node 1 be (x1,y1,z1) and that of node 2 be (x2,y2,z2)
The distance is calculated using the following Euclidean formula[3],
Distance= $[(x1-x2)^2+(y1-y2)^2+(z1-z2)^2]^{1/2}$
Once the node to node distances are calculated, the total path weight may be obtained by adding the individual node-node values.

4. GOOGLE CROWDSOURCING

Google Crowdsourcing is a phenomenon employd in the Google Traffic feature of the maps services and it displays the real-time traffic conditions of major roads and highways of over 50 countries.This works by analyzing the GPS-determined locations transferred to Google by a large number of handheld devices.Google calculates the live traffic by calculating the speed of users along the length of a road[4]. One of the commonly used methods of crowdsourcing is trilateration in which the precise location of the user is determined by the time delay in signals sent by the concerned cellular phone to three or more surrounding base stations. There is an “Opt-out” option in which the users can refrain themselves from sending their location information to Google[5]. However, it is the policy of the company to maintain the anonymity of the location and speed of the individual user.

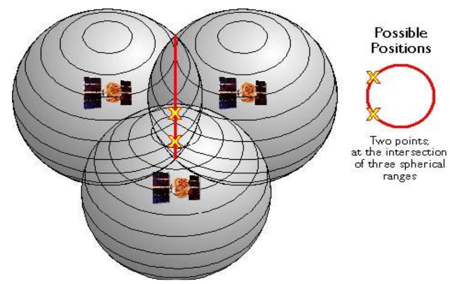

Fig 3. *Trilateration for determining the location.*

In real-time, Google Traffic represents different traffic conditions using different colors so that the commuter gets a prescient idea about the traffic.

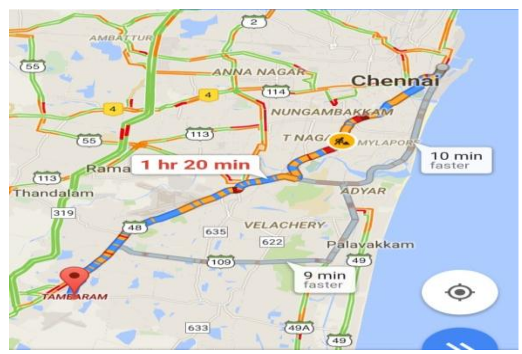

Fig 4. Real-time traffic data of Chennai city from Google Maps.

A covered overlay appears on top of major roads in the maps.The green color represents normal speed of traffic,the yellow color represents slower traffic conditions, the red color denotes congestion while a dark red color specifies nearly stopped conditions. This data is used to provide the hyperlink kind of information to calculate the Locational PageRank scheme .

5 PROPOSED ALGORITHM

5.1 Procedure

i. Calculate the various paths through which the destination may be reached from the source.
ii. The distance between each pair of successive nodes is determined from the GPS data.
iii. The total distance of each path is determined by summing the various distance values along the path[8].

iv. Provide Path Rank(PaR) for each route in the descending order,i.e the longest path gets the highest PathRank.
v. Obtain the live traffic information from the GoogleMaps which utilizes crowdsourcing technique in order to calculate the traffic density.
vi. The intermediate locations act as nodes and a mechanism similar to PageRank Algorithm is utilized to calculate the LPR for each node.
vii. The Nascent Rank(NR) is obtained by summing the PathRank along with successive values of LPR along the path.
viii. The path with the least value of NR will give the optimized route with minimal tradeoff.
ix. If two or more paths have the same NR value,the route is selected using arbitration.

5.2 Locational PageRank and Nascent Rank

These are the relatively new terms introduced in this paper. Locational PageRank is calculated using a Markov chain transition probability matrix[6]. At a particular location(node), if the road splits into two or more paths,the probability table is filled according to the traffic congestion level. The path with more denser traffic is assigned a higher probability while relatively lower density path is assigned a lower probability. In real-time maps, the path with overall higher values of red patches get the maximum probability value in the transition table followed by yellow and green patches in order. During calculation of LPR, the map is considered as bidirectional undirected graph.This is similar to the back hyperlink concept of original PageRank algorithm. Let the transition probability be X. This term is multiplied with an equal probability matrix A. Next, this AX term is further multiplied with X to obtain X^2A.Now in the next iteration this squared term is further multiplied with X to get cubed term $A(A^2X)=X^3A$.This step is repeated continuously until convergence takes palce.The converged value will yield the final LPR value. Nascent Rank is the final product of any route in this algorithm.It is calculated using the PaR value summed up with the additive LPR value from nodes covering the source to destination. **NR=PaR+∑LPR** along the path.

5.3 Arbitration

It is possible for two or more paths to get the same Nascent Rank value after calculation. This situation can be arbitrated by notifying the driver through the GPS Map.Ther are two probabilities.

- Some NRs may comprise higher values of LPR for lower value of PaR. The commuter who is willing to travel along the shortest path irrespective of the traffic conditions may choose this path.
- Some NRs may comprise lower values of LPR for higher value of PaR. The commuter who prefers to travel in a less congested path thereby trading it off with longer distances may prefer this path.

6 ILLUSTRATION

This algorithm is illustrated using a real-time traffic map of Chennai City during peak hour.

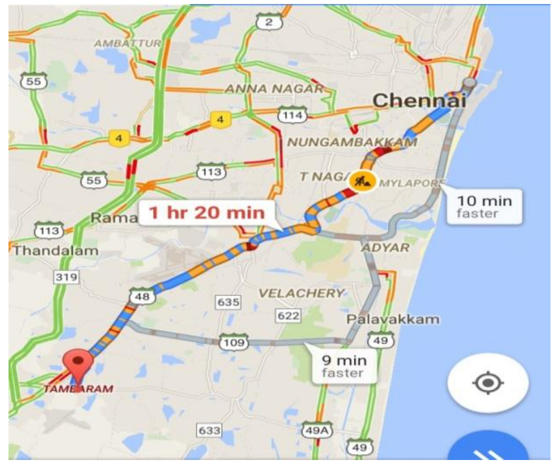

Fig 5.1. *Route 1 between Tambaram and Beach station via T Nagar.*

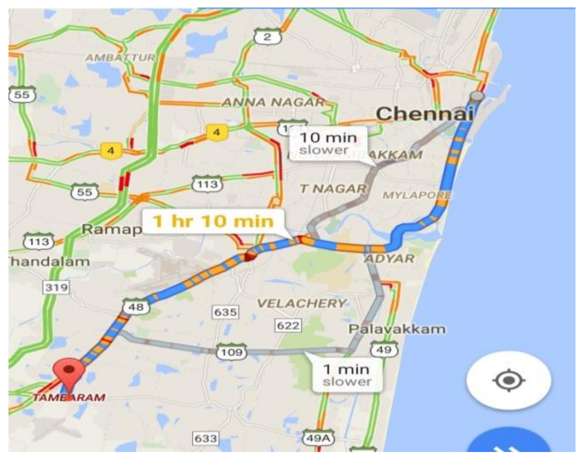

Fig 5.2. *Route 2 between Tambaram and Beach station via Adyar*

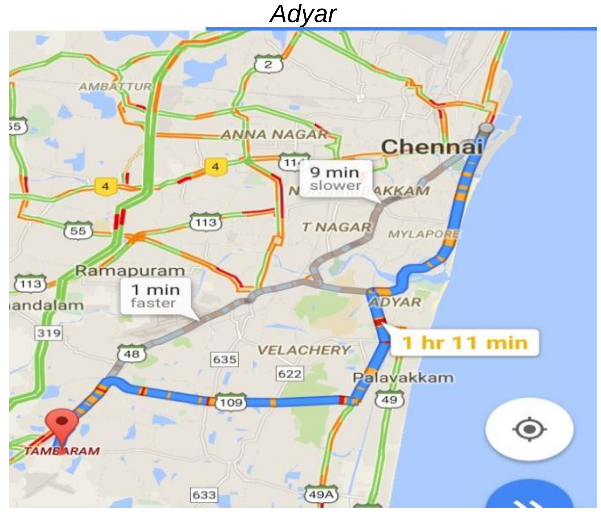

Fig 5.3. *Route 3 between Tambaram and Beach station via Mylapore.*

6.1 PathRank Calculation

PathRank is based on the total distance of each path[9] and the path with the maximum distance gets the highest rank.

TABLE 1
ROUTE 1: VIA T NAGAR

From	Till	Distance(in Km)
Tambaram	Pond's Signal	5.4
Pond's signal	Raj Bhavan	13
Raj Bhavan	Parrys	13
Parrys	Beach Station	1

Total distance: 32.4 Km

TABLE 2
ROUTE 2: VIA ADYAR

From	Till	Distance(in Km)
Tambaram	Pond's Signal	5.4
Pond's signal	Madhya Kailash	10
Madhya Kailash	Parrys	12
Parrys	Beach Station	1

Total distance: 28.4 Km

TABLE 3
ROUTE 3: VIA MYLAPORE

From	Till	Distance(in Km)
Tambaram	Pond's Signal	5.4
Pond's signal	Raj Bhavan	13
Raj Bhavan	Madhya Kailash	2.2
Madhya Kailash	Parrys	12
Parrys	Beach Station	1

Total distance: 33.6 Km

From the aforementioned data,it can be inferred that the Route 3 is the longest followed by Route 1 and Route 2.

TABLE 4
PATHRANK ASSIGNMENT

Route 1	PaR 2
Route 2	PaR 1
Route 3	PaR 3

6.2 Locational PageRank Calculation

LPR is calculated for the following map using Markov chain transition probability.

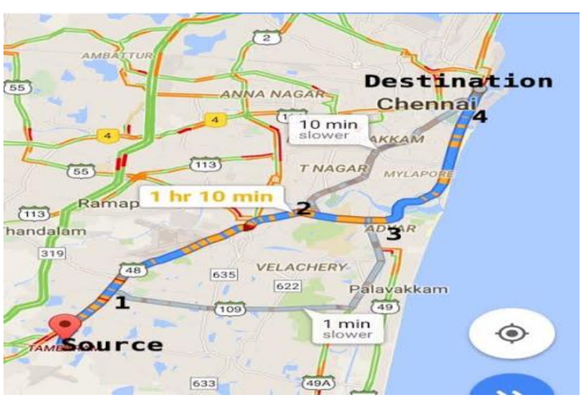

Fig 6. *Route Map with locations represented as nodes.*

From this map, we obtain two graphs. The first graph provides the information regarding nodal connections.The second graph signifies the real-time traffic flow.Although the second graph is directed,it is considered as an undirected bidirectional graph during the construction of the transition probability matrix.

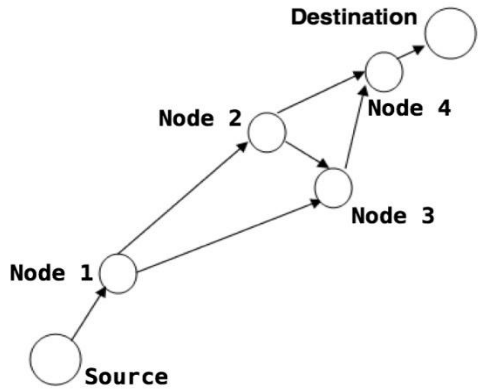

Fig 7. *Graphical representation of the map data*

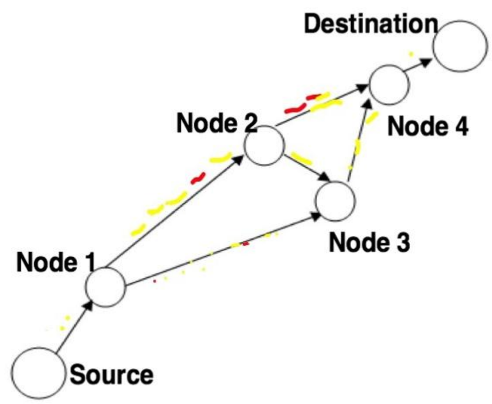

Fig 8. *Graphical representation of the map with traffic flow information.*

In the above picture, the blue data of the map is not represented.It is not included in the calculation of the transition matrix as it signifies clear traffic conditions. For the accurate estimation of the probability matrix from the color data available, image processing techniques have to be implemented in the real-time system.However,for theoretical calculations,the graphical traffic flow is sufficient.The parts of the graph with red color is assigned higher probabilities than their yellow counterparts. Markov chain transition matrix[7] is denoted by X,

$$X=\begin{bmatrix} 0 & 0.4 & 0.3 & 0 \\ 0.4 & 0 & 0.2 & 0.7 \\ 0.6 & 0.1 & 0 & 0.7 \\ 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

During the first iteration, the matrix X is multiplied with an equal probability 4x1 matrix A.The first column is assigned a probability of 0.4 for A to B and 0.6 for A to C as the path from A to C appears to be more congested than that of to B.

$$A = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

First iteration,

$$XA= \begin{bmatrix} 0.175 \\ 0.325 \\ 0.250 \\ 0.250 \end{bmatrix}$$

Second iteration,

$X(XA)=X^2A$

$$X^2A= \begin{bmatrix} 0.205 \\ 0.295 \\ 0.2125 \\ 0.2875 \end{bmatrix}$$

Third iteration,

$X(X^2A)= X^3A$

$$X^3A = \begin{bmatrix} 0.18175 \\ 0.32575 \\ 0.23875 \\ 0.25375 \end{bmatrix}$$

Fourth iteration,

$X(X^3A)= X^4A$

$$X^4A= \begin{bmatrix} 0.19405 \\ 0.31645 \\ 0.22787 \\ 0.26912 \end{bmatrix}$$

Fifth iteration,

$X(X^4A)= X^5A$

$$X^5A = \begin{bmatrix} 0.19494 \\ 0.31158 \\ 0.22881 \\ 0.27216 \end{bmatrix}$$

Sixth iteration,

$X(X^5A)= X^6A$

$$\begin{bmatrix} 0.19494 \\ 0.31158 \\ 0.22881 \\ 0.27216 \end{bmatrix}$$

The matrix converges after the sixth iteration. So this matrix gives the final probability results.

Now the LPR is assigned as per the probability.

TABLE 5
LPR DETERMINATION

Node 1	LPR 1(0.1945)
Node 2	LPR 4(0.3122)
Node 3	LPR 2(0.2283)
Node 4	LPR 3(0.2719)

6.3 Nascent Rank Calculation

For route 1,
PathRank=2
Summative LPR=1+4+3=8
As NaR=PaR+∑LPR,
Nascent Rank=2+8=10
NaR=10

For route 2,
PathRank=1
Summative LPR=1+4+2+3=10
As NaR=PaR+∑LPR,

Nascent Rank=1+10=11
NaR=11

For route 3,
PathRank=3
Summative LPR=1+2+3=6
As NaR=PaR+∑LPR,
Nascent Rank=6+3=9
NaR=9

From the data, it can be found that route 3 has the least LPR.So, in order to travel from Tambaram to Chennai Beach Station,**Route 3** via Mylapore provide the **most efficient path**,with minimal trade- off between distance need to travelled and peak hour vehicular traffic.

7 WORST CASE SCENARIO

The proposed algorithm is highly efficient only when real-time crowdsourcing data is available. This is not possible during late night hours and other occasions when the traffic flow will be extremely low. During this scenario, Locational PageRank calculation is trivial. So, it will be wise to rely on a back up algorithm which determines the shortest path to the destination neglecting the traffic congestion data. Shortest path algorithm like Dijkstra's algorithm[10] can be used to combat this situation.

8 CONCLUSION

It is mandatory to build real-time algorithms like the one proposed in this paper in order to combat various real-time situation difficulties. This algorithm , when practically implemented has the potentiality to enhance the search for optimized routing during peak hour traffic scenarios of metropolitan cities.It can be observed that the interdisciplinary approach of applying the Google PageRank algorithm of web search to real-time situation has yielded nearly optimal results.

9 REFERENCES

[1] Anatomy of a Large- scale Hypertextual Web Search by Larry Page and Sergey Brin(1997).

[2] http://www.montana.edu/gps/understd.html

[3] http://gis.stackexchange.com/questions/58530/find-Euclidean-distance-in-gps

[4] http://electronics.howstuffworks.com/how-does-google-Maps-predict-traffic.html

[5] http://www.citymetric.com/horizons/how-googles-geo-crowdsourcing-transforming-map-626

[6] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: bringing order to the web," Technical Report 1999-66, Stanford InfoLab, 1999.

[7] http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

[8] Goldberg, A.V., Harrelson, C.: Computing the shortest path: A∗ meets graph theory. In: 16th ACM-SIAM Symposium on Discrete Algorithms, pp. 156–165. ACM Press, New York (2005)

[9] Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on System Science and Cybernetics 4(2), 100–107 (1968)

[10] Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271 (1959)