Optimization of Traffic Flow of Kolkata using Centrality Measures and Load Link Analysis Techniques

Project report submitted in partial fulfillment of the requirements for the degree of Bachelor of Electronics and Telecommunication Engineering

Submitted by

AYAN BASAK
Class Roll No. - 001510701079
Exam Roll No. - ETC 198029

AISIK PAUL
Class Roll No. - 001510701087
Exam Roll No. - ETC 198046

Under the Guidance of
DR. SUDIPTA CHATTOPADHYAY
Professor
Department of Electronics and Telecommunication Engineering

JADAVPUR UNIVERSITY
KOLKATA- 700032
MAY 2019

CERTIFICATE

This is to certify that the project report entitled "**Optimization of Traffic Flow of Kolkata using Centrality Measures and Load Link Analysis Techniques"** submitted by **Ayan Basak** and **Aisik Paul** for the partial fulfillment of the degree of Bachelor of Electronics and Telecommunication Engineering of Jadavpur University is based on their assigned work during the session 2018-2019 under the guidance of **Prof. Sudipta Chattopadhyay** of **Electronics and Telecommunication Engineering Department** of **Jadavpur University,**Kolkata,West Bengal, India.

Signature of Project Guide

Signature of Student(s)

Signature of Student(s)

ACKNOWLEDGEMENT

First, we would like to thank our Project Guide, Prof. **Sudipta Chattopadhyay**, for her constant support, guidance and valuable advice during the entire course of the project work.

We would also like to thank Prof. **Sheli Sinha Chaudhuri**, the Head of the Department of Electronics and Telecommunication Engineering for her kind motivation and initiative throughout the entire course.

May 2019
Dept. of ETCE, JU, Kolkata

AYAN BASAK

AISIK PAUL

ABSTRACT

Utilizing open source temporal and spatial data from a bus tracking app that has real-time geo-tags and vital information for all private and public buses in Kolkata, to optimize traffic flow and minimize congestion. When we consider a metropolitan city like Kolkata, the road network serves as the backbone of society, controlling human and goods flow and hence, is vital to progress and prosperity. We perform a static analysis on the basis of various widely accepted centrality measures and load link analysis techniques, to find load on each stop and their importance in order to analyse and discern traffic bottlenecks. We compare these scores with real time data on congestion to evaluate their performance. The nodes can then be ranked based on the probability of being active choke points in the network. With our understanding of the network, we intend to mediate and mitigate traffic congestion. Through rerouting traffic, finding alternate bus routes, adaptive stop time to ensure a constant flow of traffic.

TABLE OF CONTENTS

Chapter 1: Introduction

1.1 OVERVIEW

This section introduces the basic background and motivation behind our work. It puts forward previous work that has been done in this area of research and specifies our objective in order to make a significant contribution to this field.

1.2 BACKGROUND

Road network is the circulatory system for modern society. So, it is important to ensure that roads are always navigable, because the alternative would be chaos. The accessibility of roads have a direct impact on both the socio-economic conditions of any place and the response time of emergency services. They greatly factor in the segregation of the entire city into residential and commercial areas while also ensuring that there is an undisturbed flow of humans and goods between them.

Cities are dynamic. With a constant influx of people, cities keep spreading, expanding, and growing, thus making the need for a robust transportation network imperative. There is a significant presence of alternate modes of transport, but they have their own limitations- ferry services are restricted to places near large water bodies, are available at specific times, only connect particular locations , and are dependent on climate conditions; rail networks are extremely expensive to build, inflexible with their timings and connecting locations, and have a high probability of being late. Therefore, the task of supporting the movement of the growing population falls on the road network. Expansion does not necessarily mean building more and bigger roads, with more vehicles on the move, it also means meticulous planning to redistribute traffic, prevent and mitigate congestions, ensuring connectivity to lesser accessible places, creating routes tailored to human flow, etc.

Every road network that we can study today is different from each other, the vary greatly in size, shape, distribution and traffic. This uniqueness can be attributed to various factors like land cover, population distribution, alternate modes of transport, kind of vehicles, etc.

Naturally, the problems attached to each network is also unique, but the one thing that remains consistent is its philosophy, to connect places and transport goods and people. So even though the graph and the linkages may change but the motivation and the ideology behind a road network remains the same.

We need to understand how congestion and inefficient road networks have an impact on society. Firstly, there is a huge economic impact. For example, Los Angeles loses a massive 19.2 billion dollars each year due to time lost on road, but the chart is topped by New York city by an even bigger g total annual loss of 33.7 billion dollars or 2982 dollars for every driver.. Although there hasn't been any Indian study replicating the same, but it is safe to assume that it will be along the same lines. The money lost due to congestion is accounted in the form of loss of productivity due to time spent stuck in traffic, higher transportation costs, fuel burnt at signals and frequent starting and stopping. Congestion also has huge impacts on the response time for various emergency services like fire fighting, first responders, ambulance services, which have serious implications on the safety and security of the city and its people alike. It also becomes a question of quality of life as a longer commute makes people more frustrated and reduces productivity.

In this case study we have considered Kolkata, a metropolitan city in east India. With a total area of 205 sq. km and a metropolitan population of 4.1 million people and a suburban population bringing the total close to 14.1 million people. Apart from the road network, it has a single line metro service, an extensive network of local trains and ferry service across the Hooghly River. With little capacity left to accommodate more people and various other factors like cost of living. and a regular major inflow of people from neighboring areas for work or for vital services like healthcare, or banking, etc. creates a strain on the entire transportation architecture of the city.

The data being used for the entire analysis comes from Pathadisha. Pathadisha is a bus tracking app, used extensively by commuters to find ETA of buses on different routes. The data parameters being recorded for each bus - location coordinates, vehicle number, plate number, direction of travel, time to next stop, bus route and speed. The data is collected from gps devices fixed in private and government busses plying on the roads of Kolkata and neighboring areas. We also have geo-locations of all the bus stops in Kolkata and surrounding areas.

What we are planning to build here is a smarter transportation network that can adjust itself according to network conditions, and is capable of preventing and mitigating congestion. We consider every bus stop as a node and every bus route as an edge with appropriate weight assigned based on the distance between the connected stops and the number of buses plying in that route. With various scoring graph network scoring algorithms we will try to understand the entire network from a quantitative point of view, and evaluate them on their effectiveness in identifying potential choke-points.

1.3 MOTIVATION

Transportation is one of the biggest problems that major cities are facing in the 21st century, more so in developing countries like India. With growing standards of living people now have the means to own personal vehicles, and easier accessibility (more local manufacturing units, easier imports, better and more efficient manufacturing techniques) have made vehicle ownership grow in huge numbers. But the transportation architecture has failed to keep up with it. Degraded road networks and inadequate flyovers have put too much pressure on the alternate means of transport that aren't and cannot be equipped to handle so many people. This need of the hour is a revamped road network architecture and state of the art planning and management systems that can react to real-time changes in network conditions.

With every major city expanding to accommodate more people and to set up commercial hubs for different industries, people are being increasingly forced to settle in the outskirts of the city owing to scarce and expensive housing and the high cost of living. These people depend heavily on the transportation system to travel to their offices and back home. Disruption and delays bear huge costs for these people, which they can't afford to pay on a daily basis. There is also a huge section of people that depend on Kolkata for resources that are not available elsewhere like advanced healthcare services, expensive brands, electronics and a wide range of goods and services. And also a daily influx of traders or businessman taking goods in and out of Kolkata. It serves as a major hub for a lot of people and places and directly impacts them. Hence the need for a robust transportation network is imperative.

1.4 LITERATURE REVIEW

Many authors have argued that the configuration of a city's street network plays an important role in traffic flow and, hence, used various centrality measures of the graph of a particular street to model and predict the flow of vehicles.

The studies of full range vulnerability have mostly been confined in principle to single link failures. Since various types of events might cover and disable extensive areas of a network, it is necessary to examine to what extent the effects differ in comparison with single link failures. Centrality measures have been employed for analysis in web networks, road networks, social networks, etc. Xing uses the Weighted Page Rank algorithm, which takes into account both the in links and out links of a node and ranks each node based on the amount of traffic. This algorithm has been proven to perform better than the conventional Page Rank algorithm. Rochat has employed closeness centrality on order to evaluate the congestion in nodes of unconnected graphs, and thus obtained a new centrality measure called harmonic centrality index. Mattsson[1], has considered multiple link failures and puts forth a full range methodology for analyzing road network vulnerability under traffic disruptions which cover an entire area. In this approach, the area of study, which includes the transport network model, is covered by cellular grids, which have uniformly shaped and sized cells; each cell represents the covering in space of an event that can cause disruption. For every cell, any links intersecting it are isolated and the effects of switching off these links are simultaneously calculated. Multiple grids which have same cell size, spaced evenly from each other, are used to enhance the accuracy of the procedure. This method permits us to systematically study the impacts of hindering events which depend on their spatial location and size.

An interesting analysis has been made by Oliveira[2], who has used the Volume/Capacity(V/C) ratio and Congestion Index(CI) to act as indicators of congestion in the network. The traffic volume is obtained by directly counting traffic or from simulations that have been created using an actual road network model. The congestion index (CI) attempts to estimate the extent of congestion in a road from by calculating the ratio between the trip time when traffic congestion is at its peak (t_c) and the time when congestion is very

low and traffic is almost flowing freely (t_f). It can be directly determined in the field, or from simulations that have been created using an actual road network model.

They have also used vulnerability indicators, where the network performance has been measured under two conditions: a) when the network is fully functional under normal conditions and is not under the influence of any external interference or before the occurrence of any sort of disturbance, and b) when the network is under the effect of a disturbance to one or more of its links.

The NRI is used as one of the vulnerability measures. It estimates the importance of a link in a road network by simulating the effect of how the network performs when that link is blocked out. Hence, a comparison is made of the total trip time or network cost under the situations that the link under analysis exists and does not exist. The NRI represents the difference in time (or cost) between the two situations, according to the formulation presented in equation:

$$NRI_k = \sum_i t_i' \times v_i' - \sum_i t_i \times v_i$$

where: NRI_k = network robustness Index for link k;

t'_i = time (or cost) of link i in the situation with link k blocked;

v'_i = traffic volume of link i in the situation with link k blocked;

t_i = time (or cost) of link i without any blocked links;

v_i = traffic volume of link i without any blocked links.

The modified network robustness index – NRI-m is similar to the NRI; it differs only in the fact that instead of completely blocking out the link whose effect on congestion is to be analyzed, the network performance is evaluated by diminishing the capacity of the link. Hence, we can calculate NRI-m for various reduced capacities.

The study of Gong[3], used a collection of intra-city trips that have been extracted from the trajectory data of taxi GPS in Shanghai to explore patterns of public commute and city structure. This study has built spatially-embedded graph networks to analyze the spatial interactions in the city and introduced network science methods to explore the sub-regional city structure based on spatial interactions of the nodes. Their approach has brought to notice a two-level hierarchical structure of Shanghai which has been examined on the basis of the duration of the taxi trips. They have then compared the administrative boundaries to the natural boundaries of the city, which have been derived from the travel patterns. In order to

increase the efficacy of urban planning, it was proposed to improve mobility and current vehicular analysis by modifying the administrative and transit planning boundaries. To achieve this objective, Shanghai was split into a 1 x 1 km cell grid, where each one of the new cells formed represented a node in the graph. The graph created was a directed graph and two nodes u and v in the graph were connected by a directed edge if there was a commute that initially originated in u and ended in v. Weights were then assigned to the edges according to the number of existing trips between the same cells. The data being analyzed was only accumulated from Monday to Thursday, as this represents a more or less constant traffic flow; it has been assumed that people indulge in increased amount of leisure and entertainment transportations near the weekends. Community analysis was then used to analyze the created network in order to point out regions within which there were common trips. Further, the detected communities were analyzed by calculating graph density, node strength, closeness centrality and betweenness centrality for each of the nodes in a community. Centers with a high degree of traffic flow were isolated using this analysis and were assigned relevant importance.

Oluwajana[4], has developed primal graphs from Akure road networks, where the edges and the nodes in the graph were described by the distance matrix. Various centrality measures were calculated for all the nodes in the network. Even though many centrality measures were able to capture different properties of the network, including some with very low values, information centrality and betweenness centrality measures were able to adapt most appropriately to the changes and the congestion variables in the network.

An interesting structural basis for analyzing taxicab data is provided by Zheng[5], in which pairs of regions (i; j) are linked to three key features: (1) number of taxis travelling from region i to region j, (2) the average speed of each taxi while travelling from region i to region j, and (3) the ratio between the actual distance of travel and the inter-centroid distance of these two regions. Zheng tries to establish flaws in current urban planning by mapping taxi trajectory data from over 20,000 taxis travelling from Beijing from April to June in 2009 and 2010 onto this framework. They try to address obvious issues in the shuttle in order to detect flaws. For example, if there is heavy traffic flow from region i to region j, but the average speed of taxicabs while commuting between these two regions is very less and a much greater distance is travelled compared to the distance between the centroids of the two regions, then one could arrive at the conclusion that there is heavy traffic flow and the alternative routes

taken are slow. Zheng tries to examine if new roads or subways have a clear impact on these problem areas by comparing and contrasting these issues over a period of two years.

Turner [6] has claimed betweenness centrality as a good predictor of vehicular flow. However, Wang[7], has criticized this approach and put forth a new model of traffic flow based on the fact that human activity is diverse and non-uniform and that the spatial interaction between two areas decreases as the distance between them increases.

Wang argues that the measure of betweenness centrality road network makes an inherent assumption that it is static and so, it cannot be used to model the actual traffic demands which is dynamic. Further, the authors have argued that traffic flow is dependent on the distance between the origin and the destination and in fact, the traffic flow decreases with trip length; the measure of betweenness centrality fails to take this factor into account. In support of their negativity, the authors compute the weighted correlation (where the weights between the nodes have been represented by the length between the streets) between "actual" traffic flow, estimated through the line-density method using a week long GPS data set of 127 vehicles in the metropolitan of Jiaozhou Bay, and the measure of betweenness centrality of the nodes of this city's street single and bidirectional networks. They establish that this measure is not accurate to predict urban traffic flow alone.

In order to examine and establish patterns of movement that exist in cities and contrast them by virtue of comparison against each other, we construct a sequence of graphs with dynamic weights which has been built from the average times of travel between sources and destinations of bus trips during various time periods of a particular day. We use centrality measures in these graphs through the entire day, considering it a dynamic model of traffic flow, and compare them to static centrality measures in a spatial graph that mimics the city's geographical structure. We use this information to identify choke points and bottlenecks in traffic and the hours during which there is maximum congestion (rush hours). Finally, we analyze how communities in our temporal networks vary through the entire day as well as how they can be compared to clusters in the static spatial network. We justify how the introduction of affordable ride sharing services such as Uber, has significantly increased the range and volume of people that use cabs as a fundamental mode of commute. Thereby using our resulting models as tools for urban planning because efficient solutions to traffic

congestion and bottlenecks can have a positive impact on the lives of a large number of people.

1.5 OBJECTIVE

The field of smart road network systems isn't something that is very new, but there have been few achievements owing to the significantly large human factor. Completely autonomous systems may not be a reality until every vehicle on the road *is* completely autonomous. But what we are trying to achieve here is a suggestion engine that will help the police maintain clear and smooth road conditions while also allowing policymakers to assess and create new routes to relieve excess stress on roads. The objectives can be divided into three distinct parts.

- Firstly we want to study the existing road network in Kolkata, considering every bus stop in the city as a node and every road joining any two bus stops without passing through another road is considered an edge. The edge maybe weighted on the basis of the distance or on the basis of the number of buses plying on it. We also try to reason which stops are more accessible and why.

- We use various node scoring algorithms to find the importance of each stop on the basis of different parameters (distance, buses, routes), and see the theoretical and practical implications of these scores. We also see which of these algorithms are suitable to be used in road network analysis.

- With our understanding of the existing road network and the various scores obtained in the previous section, we try to predict possible choke points and impact of this congestion on the entire network. We compare our results with independently collected real traffic data for accuracy.

1.6 MAJOR CONTRIBUTION

We deep dive into traffic analysis with certain centrality and load link analysis techniques commonly used in graph networks. We draw a parallel between roads and the movement of goods with a network layer moving data packets. And try to apply similar congestion mitigation techniques. With our understanding of the present network we intend to identify choke points in any arbitrary network. Help mediate traffic flow in the present network to minimize the possibility of a congestion.

1.7 ORGANIZATION

In the first section, we have provided a basic historical background of previous research work in road network congestion analysis and have also specified how we want to approach this problem. We have also stated our motivation behind approaching this relevant real-world problem. We have described the previous research work over which this problem has evolved and what has led us to work in this particular area using our approach.
The rest of the thesis is organized as follows:

Chapter 2 gives the theoretical background of our research work and also the background of research in road network congestion analysis in general.

In Chapter 3, we have described the model which we have used to perform congestion analysis and arrive at our results.

Finally, Chapter 4 presents the inference for our work and the future impact that it could have in tackling real-world congestion problems.

CHAPTER 2: THEORETICAL BACKGROUND

2.1 OVERVIEW

The congestion at each node is calculated with the help of certain measures based on the nodes and its neighbours. We need to familiarize ourselves with centrality and connectivity algorithms .

2.2 CENTRALITY MEASURES

In this section, we discuss the various centrality measures [11] we have used to evaluate congestion in the road network.

2.2.1 DEGREE CENTRALITY

Degree Centrality [16] is the simplest form of connectivity measurement and can only be used in directed graphs. The out-degree centrality for a node v is the fraction of nodes its outgoing edges are connected to . For a directed graph, $G = (V(G) , E(G))$ and a vertex $x_1 \in V(G)$,

The **Out-Degree** of x_1 refers to the number of arcs incident from x_1, that is, the number of arcs directed away from the vertex x_1.

The **In-Degree** of x_1 refers to the number of arcs incident to x_1. That is, the number of arcs directed towards the vertex x_1

$$H = \sum_{j=1}^{|Y|} [C_D(y*) - C_D(y_j)]$$

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v*) - C_D(v_i)]}{H}$$

Hence a node is considered to be important if it has many neighbors, or, in our case since we have an undirected graph, if there are a lot of roads coming out of a node (stop) and connecting to neighbouring nodes (stops). Then that node can be considered to be important. A variation of this algorithm can be performed with the help of using the number of routes that passes through the stop instead of considering just roads.

2.2.2 BETWEENNESS CENTRALITY

Betweenness centrality[14] is a measure of the influence of a vertex over the flow of information between every pair of vertices under the assumption that information primarily flows over the shortest paths between them. High centrality scores indicate that a vertex lies on a considerable fraction of shortest paths connecting pairs of vertices.

(i) Every pair of vertices in a connected graph provides a value lying in [0, 1] to the betweenness centrality[15] of all other vertices.

(ii) If there is only one geodesic joining a particular pair of vertices, then that pair provides a betweenness centrality 1 to each of its intermediate vertices and zero to all other vertices. For example, in a path graph, a pair of vertices provides a betweenness centrality 1 to each of its interior vertices and zero to the exterior vertices. A pair of adjacent vertices always provides zero to all others.

(iii) If there are geodesics of length 2 joining a pair of vertices, then that pair of vertices provides a betweenness centrality value to each of the intermediate vertices.

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

Where v is a set of nodes, the score is a fraction of the total number of shortest paths between any set of nodes passing through a particular node v over the total number of shortest paths

between any two nodes in the entire network. This is particularly useful in our scenario since people generally try to take the shortest route possible, given that we are only working with main roads big enough for a bus to travel and assuming that there is no congestion anywhere and that is not a part of the judgment process. So we can use betweenness centrality to identify nodes preferred by commuters on two and four wheelers.

2.2.3 CLOSENESS CENTRALITY

Closeness Centrality[18] is a measure of the closeness of a particular node to with respect to the entire network. A node is considered to be central if it is very easy to get to that node from every other node[19]. For a node the closeness centrality is calculated as the inverse of the total sum of shortest distances between that node and every other node in the network. This is not particularly useful for graphs that are disconnected as the distance between two disconnected nodes is infinity.

$$C(x) = \frac{1}{\sum_y d(y,x)}.$$

2.2.5 EIGENVECTOR CENTRALITY

In degree centrality we measured the importance of the node by the number of neighbours it was connected to. Eigenvector[12] is similar, but we don't consider connection to each node equally, connection to more important nodes is considered more valuable and makes the node more valuable.

2.3 RANKING ALGORITHMS

In this section, we compare the various load link algorithms that we have used in order to compare the different nodes according to their probable level of congestion and rank them accordingly.

2.3.1 PAGE RANK

Page Rank[20] is based on the idea that links from important nodes count more; that is, that importance flows across the directed edges of a graph. It is uniform apart from the outer rims of the city.

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)},$$

where L(v) represents number of out links from that vertex and PR(v) is page rank of that vertex, and PR(u) represents the updated page rank. It is an iteratively updated score.

PageRank 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. The PageRank algorithm outputs a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. It is assumed in several research papers that the distribution is evenly divided among all documents in the collection at the beginning of the computational process. The PageRank computations require several passes, called “iterations”, through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value.

Let us consider we have 4 nodes A, B, C and D. The PageRank of node A is given by the

following formula where PR ( ) is the PageRank of the respective nodes and L ( ) is the number of the outgoing links from the respective nodes.

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}.$$

This is an iterative process and at each step the PageRank is updated for each node. The terminating case is when the difference in PageRank of a particular node for two consecutive stages is smaller than the error value, set at the beginning of operation. In the first iteration each node has a PageRank of one over the total number of nodes. PageRank can be generalised in the form of:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)},$$

i.e. the PageRank value for a page u depends on the PageRank values for each page v that is present in the set Bu (the set containing all pages linking to page u), divided by the number L(v) of links from page v. The algorithm involves a damping factor for the calculation of the pagerank.

2.3.2 HITS

HITS algorithm stands for Hyperlink Induced Topic Search. In HITS algorithm nodes are divided into Hubs and Authorities. In the context of the internet it was assumed that there were some pages known as hubs which served as large directories but they weren't necessarily authorities on the information they held, but they pointed to a number of authorities. Authorities on the other end held a lot of information on the subject. We can conclude that a hub is good if it points to a large number of authorities, and an authority is good if it is pointed to by a large number of hubs.

Hubs rate nodes based on their quality as an expert and **Authorities** rate nodes based on their quality as content providers. In our temporal graphs, the authority of city regions can be interpreted as a measure of how long it takes to get to a given region. Conversely, hub scores would be interpreted as a measure of the amount of time to drive out of a city region.

$$\mathbf{hub}(p) = \sum_{i=1}^{n} \mathbf{auth}(i)$$

$$\mathbf{auth}(p) = \sum_{i=1}^{n} \mathbf{hub}(i)$$

2.4 Haversine Distance

Any point on earth is represented as a tuple of its latitude and longitude and not as cartesian coordinates. When we are trying to find the distance between any two points on earth we have to consider that it is a sphere with a radius of around 6400 kms. So we make use of the haversine formula[23].

$$\operatorname{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2}$$

$$d = r \operatorname{archav}(h) = 2r \arcsin\left(\sqrt{h}\right)$$

2.5 DBSCAN

A density based spatial clustering algorithm, DBSCAN[25] chooses a random point out of the entire dataset and checks if there are any points in its vicinity at a smaller distance than a preset value (epsilon). If so, DBSCAN considers it to be part of the cluster and continues to do so with the new points in the cluster. Else it discards the point considering it to be noise and takes up a new point. Creating these arbitrary shaped clusters for different groups. DBSCAN is density based, that is, we only set maximum distance possible between two neighbouring points.

The advantage of using DBSCAN is that we don’t have to assume the number of clusters and it is capable of separating noise and creating arbitrary shaped clusters. The disadvantage of using DBSCAN is that it doesn’t perform well when we are working with clusters of varying density.

CHAPTER 3: PROPOSED MODEL

3.1 Overview

In this section, we present a description of the model that we have used to establish traffic bottlenecks by evaluating the performance of various centrality measures

We intend to identify locations that have a possibility of being congested and propose means to mitigate it. For our purposes we draw an analogy between road network system and a network layer in communication network.
The bus is considered as the data packet and all the people inside the bus are information being transferred from one place to another. We have no control over who wants to communicate and the quantity of information, similarly we have little control over where people get up and get down. But we are regulated by the network which limits the amount of information that can flow.

A bus/packet has the following attributes:

- Has unique Start-Stop.
- Has a unique Route.
- Goal is to reach the destination.
- Speed depends on network conditions.

A road/edge has the following attributes:

- Connects any two different stops.
- Has a unique set of routes.
- Has two states: congested and clear.
- Weight between two stops is defined as connectivity.

A stop/node has the following attributes:

- Can be a terminus or a destination.
- Where the human exchange takes place.
- There maybe one or multiple routes through it.
- Importance of stop based on number of routes and centrality.

From Pathadisha we have extracted coordinates of every bus stop in and around Kolkata and their names. These shall be the nodes in our network

We need to find the route for every bus and plot all the stops it moves through and their order, so that we can connect nodes to form edges. In order to do this we track the location of every bus at an interval of 2 minutes until it starts retracing its path, we join a line along these points and find every stop they pass through. Once we have all the routes we can connect the stops and store all the routes plying through it. Now we start finding the centrality scores[20].

We start with the simplest, degree centrality. The importance of a node is decided on the basis of the number of neighbours it is directly connected to. Here we use a variation of degree centrality; we create a connection between two stops for every route that connects them instead of just on the basis of availability of road. This is a better approximation

because the ease of movement of people between two stops depends more on the number of buses plying between them than the presence of road.

Similarly we find the other centrality measures for every node in the network. With betweenness centrality we identify stops that are most likely to be used by commuters to move between any two places at random, assuming that it is equally likely for a person to start and end at any stop, we don't take demand into consideration or availability of other means of transport. A more accurate way to calculate betweenness centrality is by giving more weight to stops that have a higher demand.

With closeness centrality we assume that there is a direct relationship between the importance of a place and the density of stops. When we look at any city, places that are more crowded or frequently visited have a higher density of stops. This is primarily to help regulate traffic and allow people to get down at their desired spot.

In Eigenvector centrality we don't give equal importance to every connection, we assume that connection to a more important node is more valuable. This is particularly useful in road analysis since we have to ensure that majorly used roadways are clear even if it means there is delay in connecting roads.

HITS and PageRank are link analysis techniques frequently used in web search to find relevant nodes from the entire network. We run these on our road network to find nodes most likely to be visited when people are moving around in a random fashion.

Now moving on to the real data, we track buses at an interval of ten minutes for an entire day. For each of these frames we classify each bus as stopped, slowed down or free moving. We superimpose all the frames within an hour to separate noise (a bus that may
have stopped for some other reason) and then we use DBSCAN to cluster each of the three individual classes to identify the actual position of the congestion.
Once we have identified the congestion, we score all the nodes on the basis of their probability of being congested. We compare it with the centrality and link analysis scores we had calculated earlier and find correlation index for each. This will help us understand which of the scoring techniques might be most beneficial in identifying possible choke points.

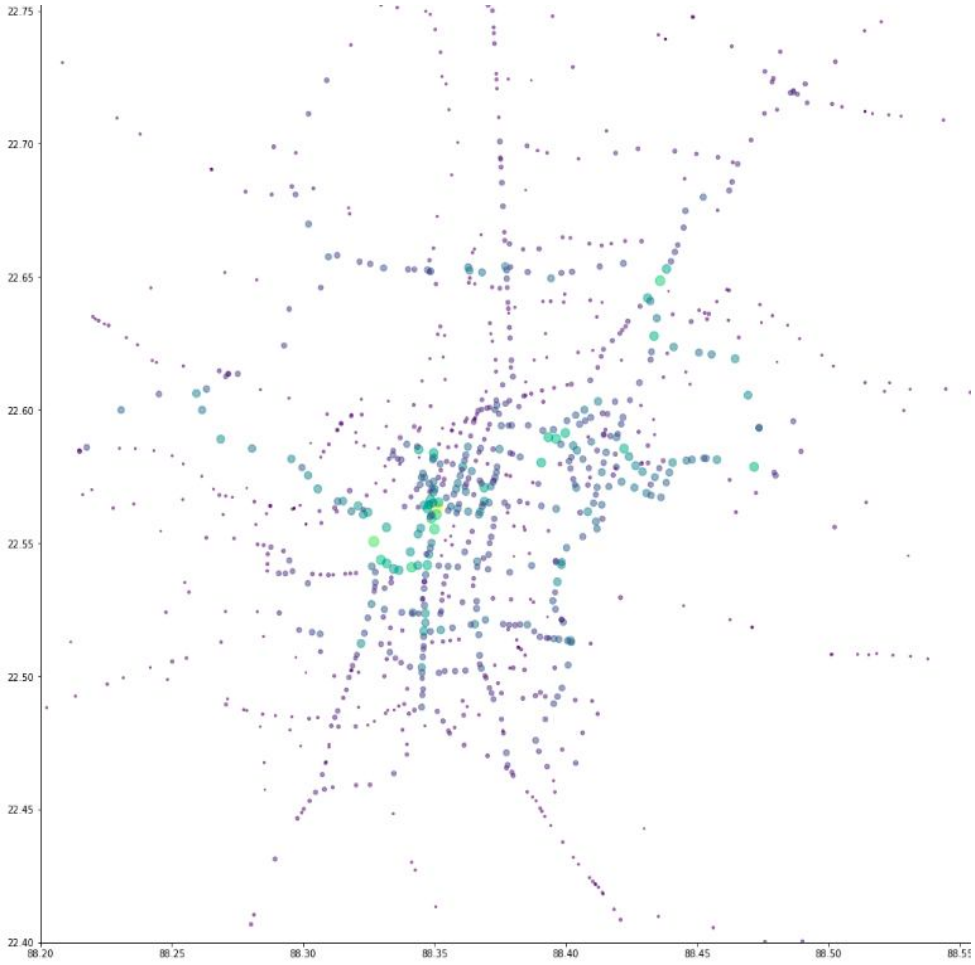
CHAPTER 4: SIMULATION RESULTS

4.1 Overview

In this section we present all our finding and try to ascertain what they mean. We aim to understand what kind of implications they may have in the perspective of a real road network. Also to understand if at all such a study is viable in identifying congestion.

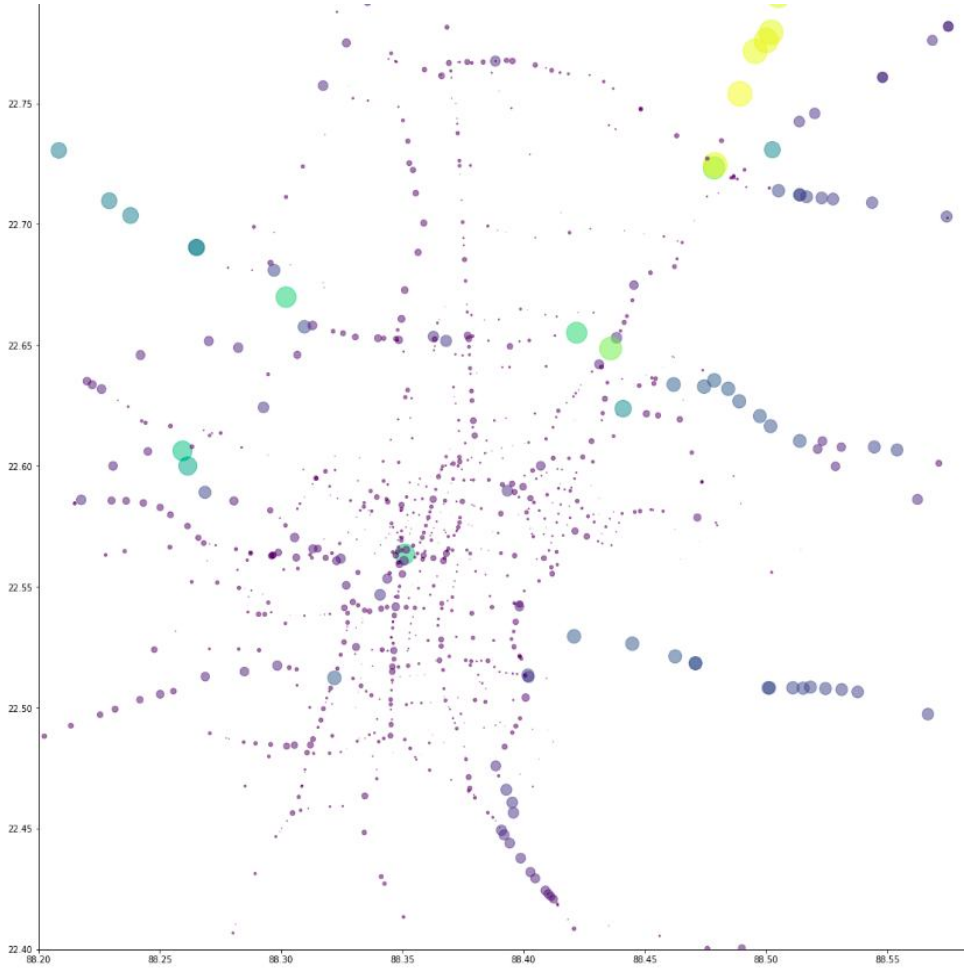
4.1 Degree Centrality

We can clearly see that places inward in Kolkata have a higher score than areas on the fringe. This is clearly justified since there is greater road connectivity and more number of bus routes in central Kolkata than in places farther outside. Since number of bus routes is an indication of the number of people moving between those stops, this is useful in understanding demand.



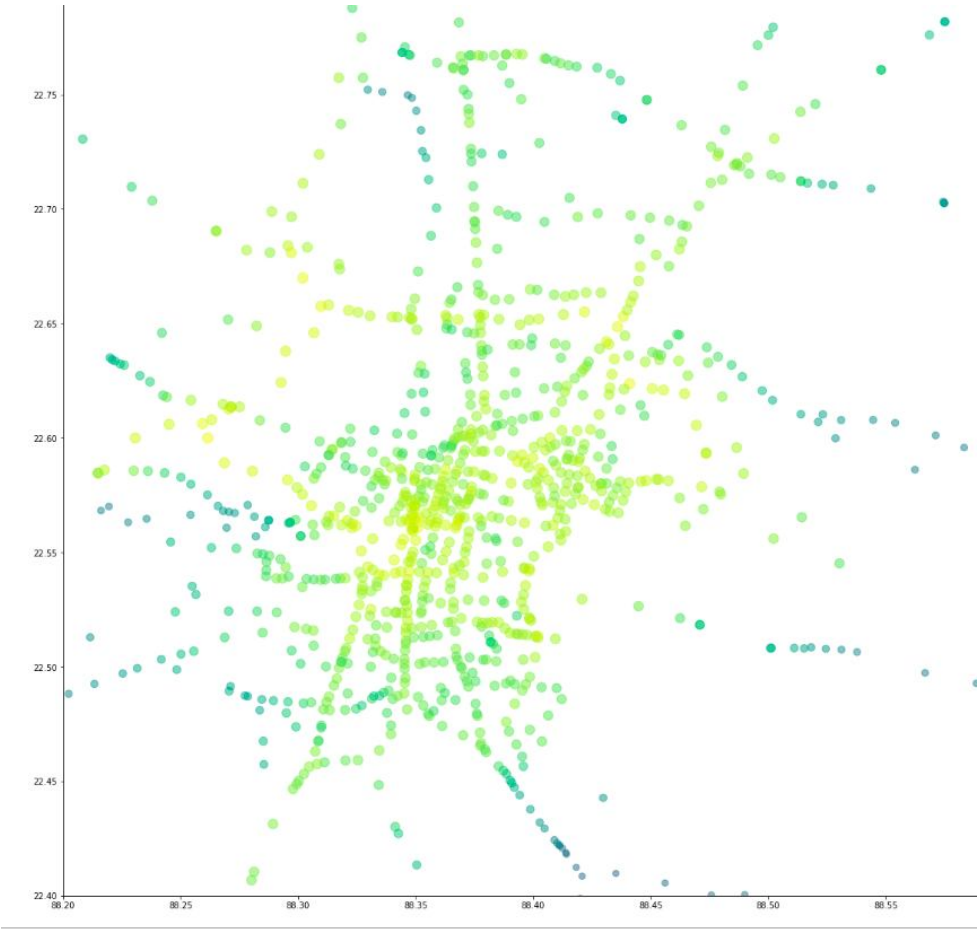
4.2 Betweenness Centrality

Betweenness centrality is useful in approximating vehicular traffic like cars and bikes since it assumes that packets move along the shortest path. But in this case we see that certain roads that go out of Kolkata have a much higher value, as seen from their bigger size and bright yellow color. This is so because we did not take demand into consideration and these roads are the only means to go in and out of Kolkata. And places at the centre like Esplanade since they must be passed whenever we move north-south or east-west.



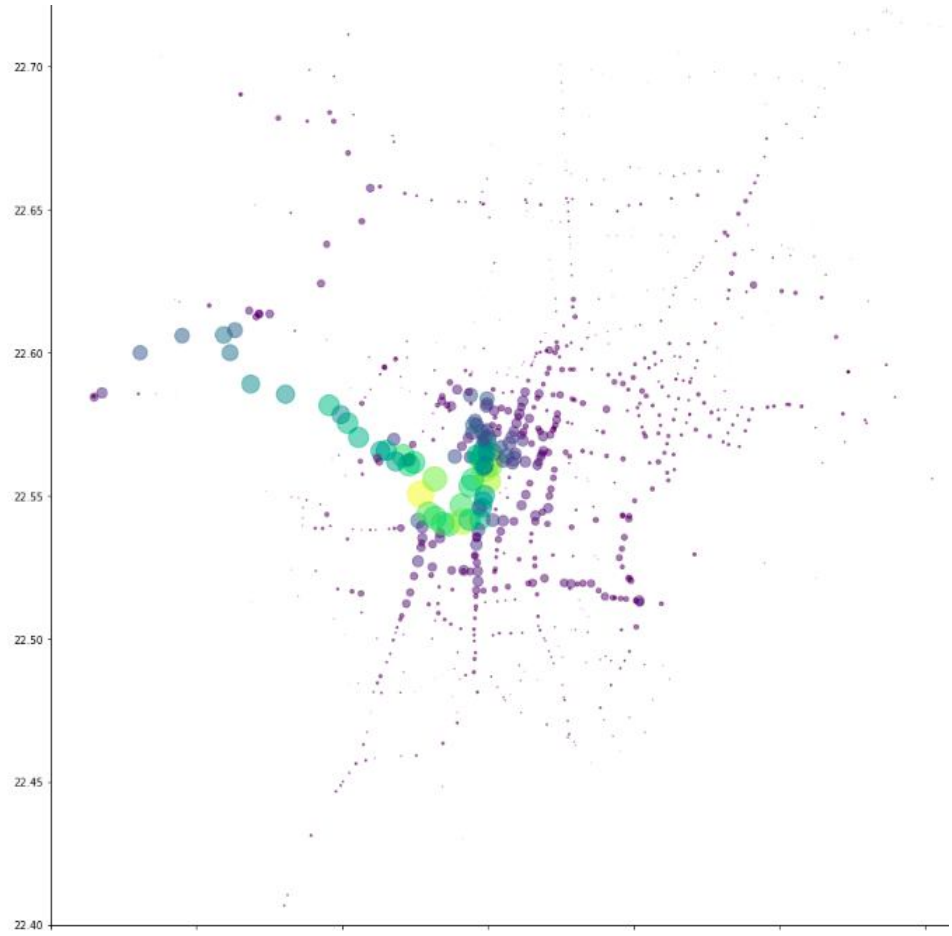
4.3 Closeness Centrality

This is particularly useful in cases with stops that are distributed non-uniformly. Since the entirety of kolkata has a pretty uniform densely packed set of stops we get a normalised value throughout, but we can separate places central to Kolkata and farther out.



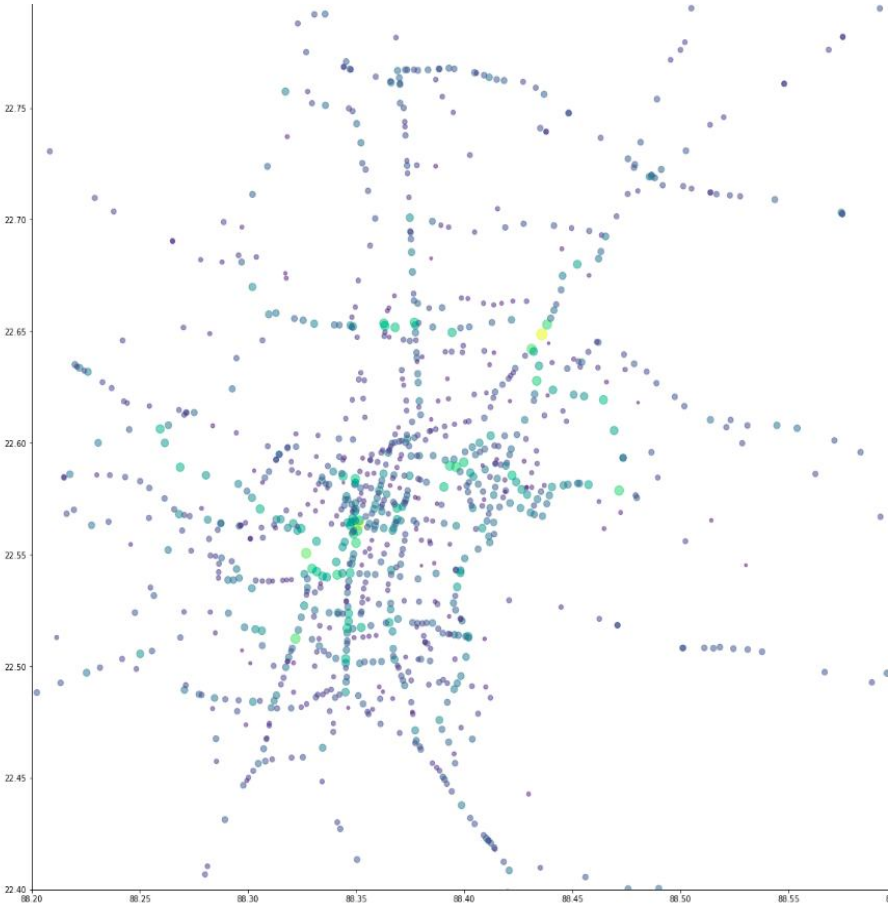
4.4 Eigenvector Centrality

In Eigenvector centrality we can see that the plot is different from the other since it accounts for the fact that places that are connected to more important nodes have greater importance. So we have a high value for places in and around esplanade like Maidan, central and on the other side of Ganga near Howrah.



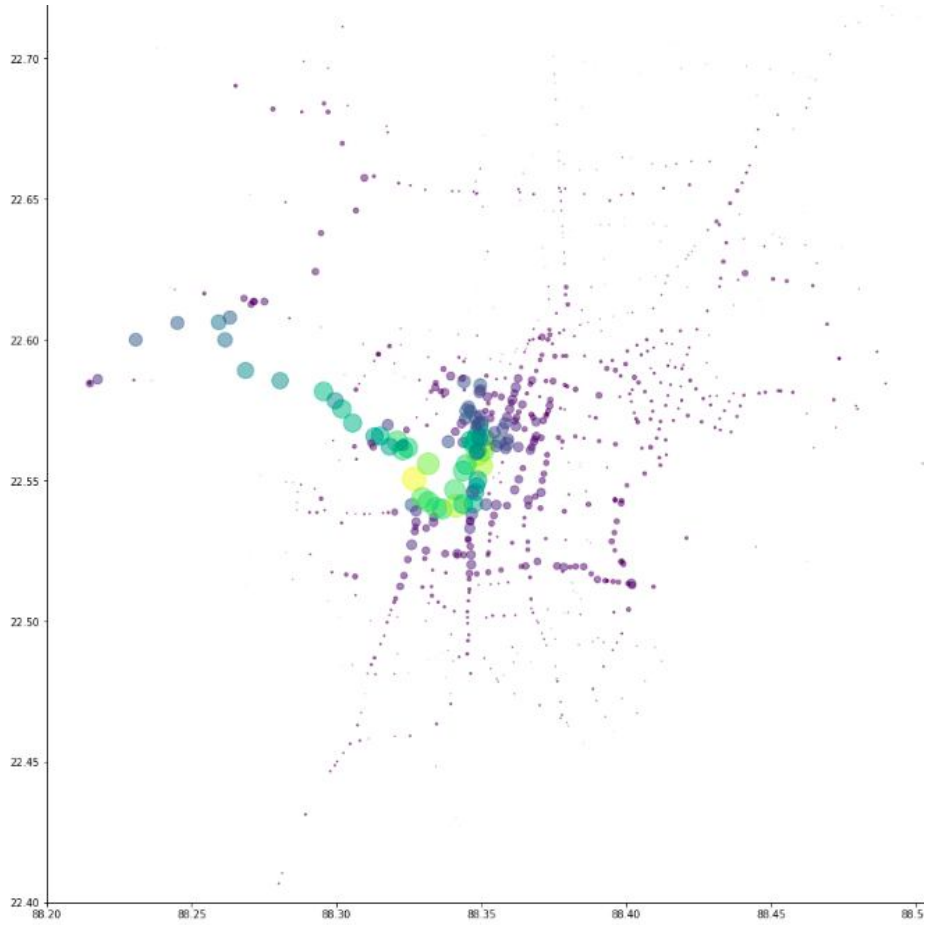
4.5 PageRank

This is useful in identifying important junctions in the entire network. As we can see the values are fairly normalized throughout, with a few bright spots like Shobha Bazar, Ultadanga, Esplanade, Ballygunge, etc. these places are important depot points which connect the rest of Kolkata with each other.



4.6 HITS

We can see that this is pretty similar to Eigenvector Centrality, this is so because they are based on a similar idea. That the connection to every node does not hold an equal value and connection to a more important node is given greater priority.

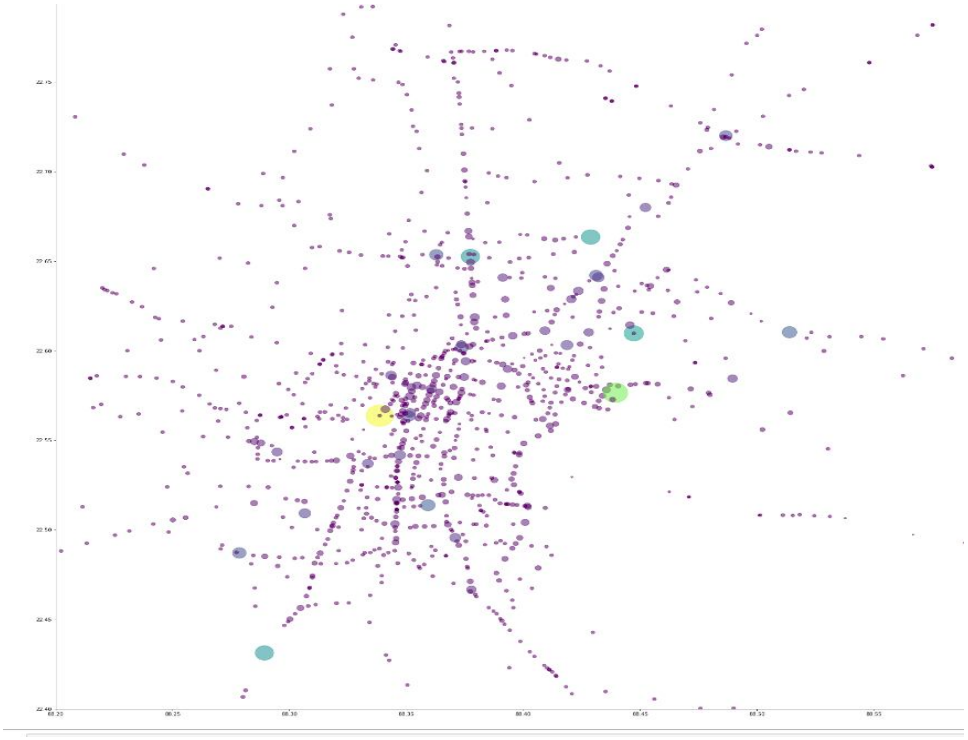


4.6 CONGESTION

After tagging all the buses for a period of 12 hours from 11am to 11pm at an interval of 10 minutes, we find a cumulative index that we can use as a probability of congestion possibility. We had tagged each bus into any of the three classes - slowed down, stopped or free moving on the basis of their speeds and the repetitive reading was to remove any noise or bias. We have only considered buses that have started their journey and have not violated their specified path. We create an relationship to enumerate congestion,

C = (r * stopped + slow) / fast

If a place has greater number of buses going through it there is a greater chance that some of them will stop, the relationship is very rudimentary, but it is useful to remove certain biases. In order to give greater importance to buses that have stopped it is multiplied by a factor r. We have arbitrarily taken ‘r’ as 2.



Even though the entire scenario changes throughout the day, we can see a similar trend throughout the day. As we saw from our centrality scores and load link analysis, places like Esplanade, Ballygunge are important as per connectivity is concerned along with Howrah, Shobha Bazar and Ultadanga. But a very important factor separates these two sets “Road Management”. Places like Howrah, Shobha Bazar and Ultadanga have to handle loads greater than their capacity, hence due to mismanagement a bus has to stop for a longer period on average. We also see a high waiting time at entry points in Kolkata.

CHAPTER 5: CONCLUSION

Our entire work is revolves around understand the present traffic architecture, we want to get an idea of the intricacies that go into creation of a congestion in order to build solutions to mitigate it. We were moved by how much of our daily life and livelihood depends on transportation, and so little thought goes into it. On an average a person spends 30 mins commuting more for people living in suburban areas, and with growing population it can only get bigger. We must act before it's too late.

With every passing year we get cars that are faster than ever. The current record stands at 305 meter in 3.64 seconds, but what's the use. The same 305 meters could take you an hour in a busy city traffic. For someone this could be losing your dream job, for someone this could be life or death. We don't need smarter cars we need smarter roads.

What we intend to propose here is a smarter, self adaptable traffic management system that can fit into the existing architecture without hindering it. An assistance system that can predict the possibility of a congestion, and suggest possible solutions to relieve it in the form of rerouting traffic, adaptive stop timings, smarter bus route planning. What we have presented here is just the tip of the iceberg.

REFERENCES

[1] Erik Jenelius and Lars-Göran Mattsson, “Road network vulnerability analysis of area-covering disruptions:A grid-based approach with case study.”

[2] Eduardo Leal de Oliveira, Licínio da Silva Portugal, Walter Porto Junior, “Determining critical links in a road network: vulnerability and congestion indicators”

[3] Xi Liu , Li Gong , Yongxi Gong ,Yongxi Gong , Yu Liu , “Revealing travel patterns and city structure with taxi trip data”

[4] Seun Daniel Oluwajana, Olufikayo Oluwaseun Aderinlewo, Adebayo Oladipo Owolabi, Silvana Vivian Croope, “Assessment of Centrality Properties of Akure Road network”

[5] Yu Zheng, Yanchi Liu1, Jing Yuan, Xing Xie, “Urban Computing with Taxicabs”

[6] Alasdair Turner, “From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis”

[7] Song Gao, Yaoli Wang, Yong Gao, and Yu Liu, “Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. Environment and Planning B: Planning and Design”

[8] Michelle Girvan and Mark EJ Newman, “Community structure in social and biological”
networks. Proceedings of the national academy of sciences

[9] Gabor Csardi and Tamas Nepusz. “The igraph software package for complex network research. InterJournal, Complex Systems”

[10] İ. Türker, “Evaluation of the Turkish Highway Network Analysis with Traffic Data”

[11] Noah E. Friedkin, “Theoretical Foundations for Centrality Measures”

[12] Phillip Bonacich, Paulette Lloyd, “Eigenvector-like measures of centrality for asymmetric relations”

[13] Paolo Crucitti , Vito Latora and Sergio Porta, “Centrality Measures in Spatial Networks of Urban Streets”

[14] Marc Barthelemy, “Betweenness Centrality in Large Complex Networks”

[15] Douglas R. White, Stephen P. Borgatti, “Betweenness centrality measures for directed graphs”

[16] Tore Opsahl, Filip Agneessens , John Skvoretz, “Node Centrality in Weighted

Networks: Generalizing Degree and Shortest Paths”

[17] Martin Everett,Stephen P. Borgatti, “Extending Centrality”

[18] Kazuya Okamoto, Wei Chen, and Xiang-Yang Li, “Ranking of Closeness Centrality for Large-Scale Social Networks”

[19] Yannick Rochat, “Closeness Centrality Extended To Unconnected Graphs : The Harmonic Centrality Index”

[20] Wenpu Xing and Ali Ghorbani, “Weighted PageRank Algorithm”

[21] Taher H. Haveliwala, “Topic-Sensitive PageRank”

[22] Fabrizio Lamberti,Andrea Sanna and Claudio Demartini, “A Relation-Based Page Rank Algorithm for Semantic Web Search Engines”

[23] Cecep Nurul Alam,Khaerul Manaf,Aldy Rialdy Atmadja,Digital Khrisna Aurum, “Implementation of Haversine Formula for Counting Event Visitor in The Radius Based on Android Application”

[24] Prof. Nitin R.Chopde1 , Mr. Mangesh K. Nichat, “Landmark Based Shortest Path Detection by Using A* and Haversine Formula”

[25] Jeffrey Erman, Martin Arlitt, Anirban Mahanti, “Traffic Classification Using Clustering Algorithms”

[26] Thanh N. Tran, Klaudia Drab, Michal Daszykowski, “Revised DBSCAN algorithm to cluster data with dense adjacent clusters”