Analyzing OSM data and Graphing using GraphX

Submitted for partial fulfilment of the Degree
of
Bachelor of Technology
(Computer Science)

Submitted By:
Sharanmeet Singh
1410926
Ranvir Singh
1410917

Submitted To:
Sukhjit Singh Sehra
Training Coordinator
CSE Department

Department of Computer Science & Engineering
Guru Nanak Dev Engineering College
Ludhiana 141006

Acknowledgement

We, students of Guru Nanak Dev Engineering College, Ludhiana, have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

The author is highly grateful to Dr. M.S. Saini Director, Guru Nanak Dev Engineering College, Ludhiana for providing her with the opportunity to carry out his Major Project in this estimed college, Guru Nanak Dev Engineering College, Ludhiana.

The author would like to whole heartedly thank Dr. Sukhjit Singh Sehra, Training Coordinator, CSE department, Guru Nanak Dev Engineering College, Ludhiana who is a vast sea of knowledge and without whose constant and never ending support and motivation, it would never have been possible to complete the project and other assignments so efficiently and effectively.

Finally, I would like to thank all the organisations and indivisuals who created such great tools and algorithms and helped us to complete the project.

Ranvir Singh

Sharanmeet Singh

Abstract

This is a research based project in which we have compared the processing time of various algorithms using apache spark and traditional serial programming on various data sets. As the size of the data sets changes, the difference in the processing time changes, producing some great results.

In this project we have used two systems one is the osmnx which is the traditional serial system in python. This uses simple serial programs to create the graphs. With small data sets the OSMnx creates the graphs faster.

On the other hand apache spark is used to create graph using the parrallel processing power of python program. With larger data sets the graph is created in lesser time.

GraphFrame is a spark API which takes data frame as input and created graphs so that we run parrallel programs using python.

Code related to this project is open source and shared using GitHub.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO ORGANISATION

Figure 1.1: Guru Nanak Dev Engineering College

Guru Nanak Dev Engineering College was established by the Nankana Sahib Education Trust Ludhiana. The Nankana Sahib Education Trust i.e NSET was founded in memory of the most sacred temple of Sri Nankana Sahib, birth place of Sri Guru Nanak Dev Ji. With the mission of Removal of Economic Backwardness through Technology Shiromani Gurudwara Parbandhak Committee i.e SGPC started a Poly technical was started in 1953 and Guru Nanak Dev Engineering College was established in 1956.

NSET resolved to uplift Rural areas by admitting 70% of students from these rural areas ever year. This commitment was made to nation on 8th April, 1956, the day foundation stone of the college building was laid by Dr. Rajendra Prasad Ji, the First President of India. The College is now ISO 9001:2000 certified.

Guru Nanak Dev Engineering College campus is spread over 88 acres of prime land about 5 Kms from Bus Stand and 8 Kms from Ludhiana Railway Station on Ludhiana-Malerkotla Road. The college campus is well planned with beautifully laid out tree plantation, pathways, flowerbeds besides

the well maintained sprawling lawns all around. It has beautiful building for College,Hostels,Swimming Pool,Sports and Gymnasium Hall Complex, Gurudwara Sahib, Bank, Dispensary, Post Office etc. There are two hostels for boys and one for girls with total accommodation of about 550 students. The main goal of this institute is:

- To build and promote teams of experts in the upcoming specialisations.
- To promote quality research and undertake research projects keeping in view their relevance to needs and requirements of technology in local industry.
- To achieve total financial independence.
- To start online transfer of knowledge in appropriate technology by means of establishing multipurpose resource centres.

CHAPTER 2

INTRODUCTION TO PROJECT

2.1 Overview

2.1.1 What is OSM?

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. The creation and growth of OSM has been motivated by restrictions on use or availability of map information across much of the world, and the advent of inexpensive portable satellite navigation devices. OSM is considered a prominent example of volunteered geographic information.

Created by Steve Coast in the UK in 2004, it was inspired by the success of Wikipedia and the predominance of proprietary map data in the UK and elsewhere. Since then, it has grown to over 2 million registered users, who can collect data using manual survey, GPS devices, aerial photography, and other free sources. This crowdsourced data is then made available under the Open Database Licence. The site is supported by the OpenStreetMap Foundation, a non-profit organisation registered in England and Wales.

Rather than the map itself, the data generated by the OpenStreetMap project is considered its primary output. The data is then available for use in both traditional applications, like its usage by Craigslist, OsmAnd, Geocaching, MapQuest Open, JMP statistical software, and Foursquare to replace Google Maps, and more unusual roles like replacing the default data included with GPS receivers. OpenStreetMap data has been favourably compared with proprietary datasources, though data quality varies worldwide.

2.1.2 Map production

Map data is collected from scratch by volunteers performing systematic ground surveys using tools such as a handheld GPS unit, a notebook, digital camera, or a voice recorder. The data is then entered into the OpenStreetMap database. Mapathon competition events are also held by OpenStreetMap team and by non-profit organisations and local governments to map a particular area.

The availability of aerial photography and other data from commercial and government sources has added important sources of data for manual editing and automated imports. Special processes

are in place to handle automated imports and avoid legal and technical problems.

2.1.3 Route planning

In February 2015, OpenStreetMap added route planning functionality to the map on its official website. The routing uses external services, namely OSRM, GraphHopper and MapQuest.

There are other routing providers and applications listed in the official Routing wiki.

2.1.4 Data storage

The OSM data primitives are stored and processed in different formats.

The main copy of the OSM data is stored in OSM's main database. The main database is a PostgreSQL database with PostGIS extension, which has one table for each data primitive, with individual objects stored as rows. All edits happen in this database, and all other formats are created from it.

For data transfer, several database dumps are created, which are available for download. The complete dump is called planet.osm. These dumps exist in two formats, one using XML and one using the Protocol Buffer Binary Format (PBF).

The LinkedGeoData data uses the GeoSPARQL and well-known text (WKT) RDF vocabularies to represent OpenStreetMap data. It is a work of the Agile Knowledge Engineering and Semantic Web (AKSW) research group at the University of Leipzig, a group mostly known for DBpedia.

2.2 User Requirement Analysis

1. Using OSMnX to get the data from OSM servers.
2. Using networkX to plot the graph with parrallel processing.
3. Using simple python program to create the graph.
4. Create the road network of an area.
5. Compare the processing times of the parrallel vs serial processors.

2.3 Feasibility Study

This study is made to see if the project on completion will serve the purpose of the organization for the amount of work, effort and the time that spend on it. Feasibility study lets the developer foresee the future of the project and the usefulness.
A feasibility study of a system proposal is according to its workability, which is the impact on the organization, ability to meet their user needs and effective use of resources. Carrying out a feasibility study involves information assessment, information collection and report writing. The information assessment phase identifies the information that is required to answer the three questions set out

above.
Once the information has been identified, you should question information sources to discover the answers to these questions Thus when a new application is proposed it normally goes through a feasibility study before it is approved for development.

A feasibility study is designed to provide an overview of the primary issues related to a business idea. The purpose is to identify any make or break issues that would prevent your business from being successful in the marketplace. In other words, a feasibility study determines whether the business idea makes sense. A thorough feasibility analysis provides a lot of information necessary for the business plan. For example, a good market analysis is necessary in order to determine the project's feasibility. This information provides the basis for the market section of the business plan.
The objective of the feasibility study is to establish the reasons for developing the software that is acceptable to users, adaptable to change and conformable to established standards.
Objectives of feasibility study are listed below:

- To analyze whether the software will meet organizational requirements.
- To determine whether the software can be implemented using the current technology and within the specified budget and schedule.
- To determine whether the software can be integrated with other existing software.

2.4 Types of Feasibility

2.4.1 Technical Feasibility

Technical feasibility is one of the first studies that must be conducted after the project has been selected. The main objective is to make sure all the technical requirements shold be analyzedd and made sure proper technologies are vailable and wisely chosn to make sure the project reaches its desired conclusion. The following should be taken to consideration:

- Technologies are searched and Graph analysis tools are available and chosen accordingly.
- The Technologies can be implemented with given resources.
- The human and economic factor is not a problem.
- The problem is to analyze the technologies available and chose wisely.

The system must be evaluated from the technical point of view first. The assessment of this feasibility must be based on an outline design of the system requirement in the terms of input, output, programs and procedures. Having identified an outline system, the investigation must go on to suggest the type of equipment, required method developing the system, of running the system once it has been designed. Technical feasibility assesses the current resources (such as hardware and software) and technology, which are required to accomplish user requirements in the software within the allocated time and budget. For this, the software development team ascertains whether the current resources and technology can be upgraded or added in the software to accomplish specified user requirements. A Technical feasibility also performs the following tasks.

- Analyzes the technical skills and capabilities of the software development team members, In this case technical skills are available with the team members.
- Determines whether the relevant technology is stable and established, In this case technology used is Networkx and GraphFrames.
- Ascertains that the technology chosen for software development has a large number of users so that they can be consulted when problems arise or improvements are requisred, much needed support is availble online in this case.

s

Technical issues raised during the investigation are:

- Does the technologies chosen can meet the requirements of task to be fullfiled?, yes the technologies are more than capable.
- Can the system expand if developed?, scalabilty is the main feature of GraphFrames Technology.

The project should be developed such that the necessary functions and performance are achieved within the constraints. The project is developed within latest technology. Through the technology may become obsolete after some period of time, due to the fact that never version of same software supports older versions, the system may still be used. So there are minimal constraints involved with this project. The system has been developed using PHP the project is technically feasible for development.

2.4.2 Economic Feasibility

The purpose of the economic feasibility assessment is to determine the positive economic benefits to the organization that the proposed system will provide. It includes quantification and identification of all the benefits expected. This assessment typically involves a cost/ benefits analysis.

Economic feasibility is the cost and logistical outlook for a business project or endeavor. Prior to embarking on a new venture, most businesses conduct an economic feasibility study, which is a study that analyzes data to determine whether the cost of the prospective new venture will ultimately be profitable to the company. Economic feasibility is sometimes determined within an organization, while other times companies hire an external company that specializes in conducting economic feasibility studies for them.
The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require. Economic feasibility determines whether the required software is capable of generating financial gains for an organization. In addition, it is necessary to consider the benefits that can be achieved by developing the software. Software is said to be economically feasible if it focuses on the issues listed below.

- Cost incurred on software development to produce long-term gains for an organization.
- Cost required to conduct full software investigation (such as requirements elicitation and requirements analysis).
- Cost of hardware, software, development team, and training.

The following are some of the important financial conclusions are made during preliminary investigation:

- The costs and economic constraints won't be a problem.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Economic analysis is the most frequently used method to determine the cost/benefit factor for evalu- ating the effectiveness of a new system. In this analysis we determine whether the benefit is gained according to the cost invested to develop the project or not. If benefits outweigh costs, only then the decision is made to design and implement the system. It is important to identify cost and benefit factors, which can be categorized as follows:

- Development Cost
- Operation Cost

This System is Economically feasible with 0 Development and Operating Charges as it is developed in Qt Framework and Octave which is open source technology and is available free of cost on the internet.

2.4.3 Operational Feasibility

Operational feasibility is a measure of how well a project solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. All the operations performed in the software are very quick and satisfy all the requirements.

2.4.4 Technological Feasibility

Technological feasibility is carried out to determine whether the project has the capability, in terms of software, hardware, personnel to handle and fulfill the user requirements. The assessment is based on an outline design of system requirements in terms of Input, Processes, Output and Procedures. Automated Building Drawings is technically feasible as it is built up using various open source technologies and it can run on any platform.

2.4.5 Behavioral Feasibility

Behavioral feasibility assesses the extent to which the required software performs a series of steps to solve business problems and user requirements. It is a measure of how well the solution of problems or a specific alternative solution will work in the organization. It is also measure of how people feel about the system. If the system is not easy to operate, than operational process would be difficult. The operator of the system should be given proper training. The system should be made such that the user can interface the system without any problem.

Operational feasibility is a measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. The operational feasibility assessment focuses on the degree to which the proposed development projects fits in with the existing business environment and objectives with regard to development schedule, delivery date,corporate culture, and existing business processes.

To ensure success, desired operational outcomes must be imparted during design and development. These include such design-dependent parameters such as reliability, maintainability, supportability, usability, producibility, disposability, sustainability, affordability and others. These parameters are required to be considered at the early stages of design if desired operational behaviors are to be realized. A system design and development requires appropriate and timely application of engineering and management efforts to meet the previously mentioned parameters. A system may serve its intended purpose most effectively when its technical and operating characteristics are engineered into the design. Therefore, operational feasibility is a critical aspect of systems engineering that needs to be an integral part of the early design phasesThis feasibility is dependent on human resources (software development team) and involves visualizing whether the software will operate after it is developed and be operative once it is installed. Operational feasibility also performs the following tasks.

- Determines whether the problems anticipated in user requirements are of high priority.
- Determines whether the solution suggested by the software development team is acceptable.
- Analyzes whether users will adapt to a new software.
- Determines whether the organization is satisfied by the alternative solutions proposed by the software development team.

Following conclusions are made after analysis:

- Enough support is available for GraphFrames and NetworkX.
- No harm is caused in development of any kind.
- The project would be beneficial because it satisfies the objectives when developed and installed. All behavioral aspects are considered carefully and conclude that the project is behaviorally feasible.

2.5 Objective of Project

The main objective of the project is to compare the processing time of graph creation using both parrallel and serial processing systems. Subobjectives of the project are:

1. Using OSMnx to get the data from OSM servers.
2. Using NetworkX to create different graphs using different algorithms.
3. Using centrality degree algorithm to plot the graph.
4. Using simple python program to create graph.
5. Comparing the processing time in both cases on different sizes of datasets.

CHAPTER 3

PROJECT DESIGN

3.1 Software Requirement Analysis

Software requirement analysis is a process of gathering and interpreting facts, diagnosing problems and the information to recommend improvements on the system. It is a problem solving activity that requires intensive communication between the system users and system developers. System analysis or study is an important phase of any system development process. The system is studied to the minutest detail and analyzed. The system analyst plays the role of the interrogator and dwells deep into the working of the present system. The system is viewed as a whole and the input to the system are identified. The outputs from the organizations are traced to the various processes. System analysis is concerned with becoming aware of the problem, identifying the relevant and decisional variables, analyzing and synthesizing the various factors and determining an optimal or at least a satisfactory solution or program of action.

A detailed study of the process must be made by various techniques like interviews, questionnaires etc. The data collected by these sources must be scrutinized to arrive to a conclusion. The conclusion is an understanding of how the system functions. This system is called the existing system. Now the existing system is subjected to close study and problem areas are identified. The designer now functions as a problem solver and tries to sort out the difficulties that the enterprise faces. The solutions are given as proposals. The proposal is then weighed with the existing system analytically and the best one is selected. The proposal is presented to the user for an endorsement by the user. The proposal is reviewed on user request and suitable changes are made. This is loop that ends as soon as the user is satisfied with proposal.

Preliminary study is the process of gathering and interpreting facts, using the information for further studies on the system. Preliminary study is problem solving activity that requires intensive communication between the system users and system developers. It does various feasibility studies. In these studies a rough figure of the system activities can be obtained, from which the decision about the strategies to be followed for effective system study and analysis can be taken.

3.1.1 Functional Requirements

- **Specific Requirements**: This phase covers the whole requirements for the system. After

understanding the system we need the input data to the system then we watch the output and determine whether the output from the system is according to our requirements or not. So what we have to input and then what we'll get as output is given in this phase. This phase also describe the software and non-function requirements of the system.

- **Input Requirements of the System**

  1. Data set of an area.
  2. Type of graph to be generated.
  3. Whether to use parrallel or serial processing.

- **Output Requirements of the System**

  1. Final time computation after processes
  2. Results from the analysis queries.

- **Software Requirements**

  1. Programming language: Python 2.7+
  2. software: LaTeX
  3. Processing Tchnology: GraphFrame
  4. Text Editor: Vim
  5. Operating System: Ubuntu 14.04+
  6. Revision System: Git

3.2 Other Specifications

A Software Requirements Analysis for a software system is a complete description of the behavior of a system to be developed. It include functional Requirements and Software Requirements. In addition to these, the SRS contains non-functional requirements. Non-functional requirements are requirements which impose constraints on the design or implementation.

- **Purpose**: To compare the processing time of serial and parrallel processing using graph generation of large data set. Perform most of difficult Calculation work.

3.3 Problem Formulation

Comparing the processing time of graph generation using parrallel as well as the serial python programming. For parrallel processing we will be using the combination of tools like NetworkX, OSMnx and GraphFrames.
When analytical solution is impossible, which was discussed by eg. Alexander Sadovsky. This means that we have to apply numerical methods in order to find the solution. This does not define that we must do calculations with computer although it usually happens so because of the number of required operations.

3.4 Facilities required for proposed work

3.4.1 Hardware Requirements

- Operating System: Linux
- Processor Speed: 512KHz or more
- RAM: Minimum 1GB

3.4.2 Software Requirements

- Softwares: NetworkX, OSMnx, GraphFrame
- Programming Language: Python 2.7+

3.5 Methodology

- Using OSMnx to get the data from OSM servers.
- Using NetworkX to create different graphs using different algorithms.
- Performing Analysis on graphs.
- Using Apache Spark via GraphFrames API to generate graphs and analyze them.
- Comparing the processing time in both cases on different sizes of datasets.

3.6 Project Work

Studied Previous System:
Before starting the project.

Learn the usage of various softwares:
Before starting with project, we have to go through the basics of tools like OSMnx, NetworkX and GraphFrame. We also have to study about the various formats in which data is accepted by the tools.

Get Familiar with Different methods and their algorithms:

Once, we have gone through algorithms of these softwares and tools, the implementation becomes easy.

Input:
Input values are taken from user or default values defined in the file are used.

Output:
The iterations are performed and processing times are computed.

CHAPTER 4

DEVELOPMENT AND IMPLEMENTATION

4.1 Introduction to tools

These are the tools that help us to reduce the work of the developer by just providing the function ready for the direct usage.

4.1.1 OSMnx

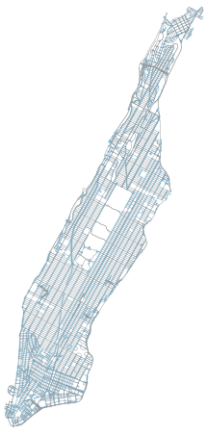

Figure 4.1: OSMnx map of manhattan

OSMnx: retrieve, construct, analyze, and visualize street networks from OpenStreetMap. OSMnx is a Python package that lets you download spatial geometries and construct, project, visualize, and analyze street networks from OpenStreetMaps APIs. Users can download and construct walkable, drivable, or bikable urban networks with a single line of Python code, and then easily analyze and visualize them.

Features

- Download street networks anywhere in the world with a single line of code
- Download other infrastructure network types, place polygons, or building footprints as well

- Download by city name, polygon, bounding box, or point/address + network distance
- Get drivable, walkable, bikable, or all street networks
- Visualize the street network as a static image or leaflet web map
- Simplify and correct the networks topology to clean and consolidate intersections
- Save networks to disk as shapefiles or GraphML
- Conduct topological and spatial analyses to automatically calculate dozens of indicators
- Calculate and plot shortest-path routes as a static image or leaflet web map
- Plot figure-ground diagrams of street networks and/or building footprints
- Download node elevations and calculate edge grades
- Visualize travel distance and travel time with isoline and isochrone maps
- Calculate and visualize street bearings and orientations

Installation

```
$ sudo apt-get install python-pip python-virtualenv
$ virtualenv venv
$ source venv/bin/activate
$ pip install osmnx
```

Usage

```
import osmnx as ox
G = ox.graph_from_place('Punjab, India', network_type='drive')
ox.plot_graph(ox.project_graph(G))
```

4.1.2 NetworkX

Figure 4.2: NetworkX logo

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Features

- Data structures for graphs, digraphs, and multigraphs

- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be ”anything” (e.g., text, images, XML records)
- Edges can hold arbitrary data (e.g., weights, time-series)
- Open source 3-clause BSD license
- Well tested with over 90% code coverage
- Additional benefits from Python include fast prototyping, easy to teach, and multi-platform

Installation

```
$ sudo apt-get install python-pip python-virtualenv
$ virtualenv venv
$ source venv/bin/activate
$ pip install networkx
```

Algorithm PageRank computes a ranking of the nodes in the graph G based on the structure of the incoming links. It was originally designed as an algorithm to rank web pages.

Graph types

- Undirected Simple
- Directed Simple
- With Self-loops
- With Parallel edges

4.1.3 GraphFrames

GraphFrames is a package for Apache Spark which provides DataFrame-based Graphs. It provides high-level APIs in Scala, Java, and Python. It aims to provide both the functionality of GraphX and extended functionality taking advantage of Spark DataFrames. This extended functionality includes motif finding, DataFrame-based serialization, and highly expressive graph queries.

What are GraphFrames?
GraphX is to RDDs as GraphFrames are to DataFrames.

GraphFrames represent graphs: vertices (e.g., users) and edges (e.g., relationships between users). If you are familiar with GraphX, then GraphFrames will be easy to learn. The key difference is that GraphFrames are based upon Spark DataFrames, rather than RDDs.

GraphFrames also provide powerful tools for running queries and standard graph algorithms. With GraphFrames, you can easily search for patterns within graphs, find important vertices, and

more. Refer to the User Guide for a full list of queries and algorithms.

creating nodes using pagerank algorithm

```
# Create a Vertex DataFrame with unique ID column "id"
v = sqlContext.createDataFrame([
  ("a", "Alice", 34),
  ("b", "Bob", 36),
  ("c", "Charlie", 30),
], ["id", "name", "age"])

# Create an Edge DataFrame with "src" and "dst" columns
e = sqlContext.createDataFrame([
  ("a", "b", "friend"),
  ("b", "c", "follow"),
  ("c", "b", "follow"),
], ["src", "dst", "relationship"])

# Create a GraphFrame
from graphframes import *
g = GraphFrame(v, e)

# Query: Get in-degree of each vertex.
g.inDegrees.show()

# Query: Count the number of "follow" connections in the graph.
g.edges.filter("relationship = 'follow'").count()

# Run PageRank algorithm, and show results.
results = g.pageRank(resetProbability=0.01, maxIter=20)
results.vertices.select("id", "pagerank").show()
```

4.2 Introduction to OSM

Figure 4.3: OSM foundation

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. The creation and growth of OSM has been motivated by restrictions on use or availability of map information across much of the world, and the advent of inexpensive portable satellite navigation devices. OSM is considered a prominent example of volunteered geographic information.

Created by Steve Coast in the UK in 2004, it was inspired by the success of Wikipedia and the predominance of proprietary map data in the UK and elsewhere. Since then, it has grown to over 2 million registered users, who can collect data using manual survey, GPS devices, aerial photography, and other free sources. This crowdsourced data is then made available under the Open Database Licence. The site is supported by the OpenStreetMap Foundation, a non-profit organisation registered in England and Wales.

Rather than the map itself, the data generated by the OpenStreetMap project is considered its primary output. The data is then available for use in both traditional applications, like its usage by Craigslist, OsmAnd, Geocaching, MapQuest Open, JMP statistical software, and Foursquare to replace Google Maps, and more unusual roles like replacing the default data included with GPS receivers. OpenStreetMap data has been favourably compared with proprietary datasources, though data quality varies worldwide.

Map usage Map is available on the following platform.

- **Web browser** Data provided by the OpenStreetMap project can be viewed in a web browser with JavaScript support via Hypertext Transfer Protocol (HTTP) on its official website.

- **OsmAnd** OsmAnd is free software for Android and iOS mobile devices that can use offline vector data from OSM. It also supports layering OSM vector data with prerendered raster map tiles from OpenStreetMap and other sources.

- **Maps.me** Maps.me is free software for Android and iOS mobile devices that provides offline maps based on OSM data.

- **GNOME Maps** GNOME Maps is a graphical front-end written in JavaScript and introduced in GNOME 3.10. It provides a mechanism to find the user's location with the help of GeoClue, finds directions via GraphHopper and it can deliver a list as answer to queries.

- **Marble** Marble is a KDE virtual globe application which received support for OpenStreetMap.

- **FoxtrotGPS** FoxtrotGPS is a GTK+-based map viewer, that is especially suited to touch input. It is available in the SHR or Debian repositories.

- **Emerillon** Another GTK+-based map viewer.

- The web site OpenStreetMap.org provides a slippy map interface based on the Leaflet JavaScript library (and formerly built on OpenLayers), displaying map tiles rendered by the Mapnik rendering engine, and tiles from other sources including OpenCycleMap.org.

- Custom maps can also be generated from OSM data through various software including Jawg Maps, Mapnik, Mapbox Studio, Mapzen's Tangrams.

- OpenStreetMap maintains lists of online and offline routing engines available, such as the Open Source Routing Machine. OSM data is popular with routing researchers, and is also available to open-source projects and companies to build routing applications (or for any other purpose).

4.3 Ubuntu: An open source OS

Figure 4.4: Ubuntu

During my training, I also got familiar with a great and open source Operating System, Ubuntu. Firstly, it was quite difficult for a regular MS Windows user to port to Ubuntu. I did all of my project work using this vast operating system.
Ubuntu is a Debian-based Linux operating system, with Unity as its default desktop environment. It is based on free software and named after the Southern African philosophy of ubuntu (literally, ”human-ness”), which often is translated as ”humanity towards others” or ”the belief in a universal bond of sharing that connects all humanity”.
It came under Linux. A kernel normally used by many of the computer persons.You will rarely see a person who is unaware of the term Linux. From perspective of a computer simpleton the one who uses linux mostly shall be having a good knowledge regarding the working of shell,kernel etc.

Linux was created by Linus Torvalds. One of a gem of computer scientist who is popular for his OS.
Linus was one of the student in Finland and had read a book Operating Systems: Design and Implementation by Andrew S. Tanenbaum.
In this book the professor explained about the working of Kernel. To my strange is that he had given the whole source code of his kernel named MINIX in that book. Its really weird but acts as a lucky draw for LInus who took interest in this and with the help of MINIX he created a new OS named Linux. He had told about Andrew in his acknowledgement.

Ubuntu’s goal is to be secure ”out-of-the box”. By default user’s programs run with low privileges and cannot corrupt the operating system or other user’s files. For increased security, the sudo tool is used to assign temporary privileges for performing administrative tasks, which allows the root account to remain locked and helps prevent inexperienced users from inadvertently making catastrophic system changes or opening security holes.

4.4 Introduction to Reveal-js & Reveal-md

Figure 4.5: MD & JS

Reveal-js is one of the framework of Javascript. This can be used for presentations purpose.
Now before going to reveal-md lets talk about some fundamental things.
What is a Markup language?
Markup languages are designed for the processing, definition and presentation of text. The language specifies code for formatting, both the layout and style, within a text file. HTML and Markdown is an example of a widely known and used markup language.
Markdown is a lightweight Markup Language with simple plain text formatting syntax designed so that it can be converted to HTML and many other formats. It is created by John Gruber. It had '.md' or '.markdown' extention.
"Markdown is a text-to-HTML conversion software tool written in Perl for web writers."
Moreover, to enable markdown feature of reveal.js, we need reveal-md. The Markdown feature of reveal.js is awesome, and has an easy (and configurable) syntax to separate slides. Use three dashes surrounded by two blank lines.

4.4.1 Installation of reveal-md

Installation of reveal-md is a very easy process. Type the commands in the terminal:

$ sudo apt-get install npm

$ sudo apt-get install nodejs-legacy

$ sudo npm install -g reveal-md

This will install reveal-md on your pc or laptop.

4.5 Introduction to LaTeX

LaTeX, I had never heard about this term before doing this project, but when I came to know about it's features, found it excellent. LaTeX is a document markup language and document preparation system for the TeX typesetting program. Within the typesetting system, its name is styled as LaTeX.

Figure 4.6: Donald Knuth, Inventor Of TeX typesetting system

Within the typesetting system, its name is styled as LaTeX. The term LaTeX refers only to the language in which documents are written, not to the editor used to write those documents. In order to create a document in LaTeX, a .tex file must be created using some form of text editor. While most text editors can be used to create a LaTeX document, a number of editors have been created specifically for working with LaTeX.

LaTeX is most widely used by mathematicians, scientists, engineers, philosophers, linguists, economists and other scholars in academia. As a primary or intermediate format, e.g., translating DocBook and other XML-based formats to PDF, LaTeX is used because of the high quality of typesetting achievable by TeX. The typesetting system offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

LaTeX is intended to provide a high-level language that accesses the power of TeX. LaTeX essentially comprises a collection of TeX macros and a program to process LaTeXdocuments. Because the TeX formatting commands are very low-level, it is usually much simpler for end-users to use LaTeX.

To run LATEX on your own computer, you need to use a latex distribution. A distribution includes a latex program and (typically) several thousand packages.

- On Windows: MikTEX or TEXLive
- On Linux: TEXLive
- On Mac: MacTEX

4.5.1 Typesetting

LaTeXwas first developed in 1985 by Leslie Lamport. In preparing a LaTeX document, the author specifies the logical structure using familiar concepts such as chapter, section, table, figure, etc., and lets the LaTeX system worry about the presentation of these structures. It therefore encourages the separation of layout from content while still allowing manual typesetting adjustments where needed.

```
\documentclass[12pt]{article}
\usepackage{amsmath}
\title{\LaTeX}
\date{}
\begin{document}
  \maketitle
  \LaTeX{} is a document preparation system
  for the \TeX{} typesetting program.
\end{document}
```

Apart from this lat.pdf; lat.aux, lat.log, lat.pdf files are created by default.

- AUX is a data file format used by Latex AUX is a data file format used by LaTeX. LaTeX is a macro package which uses TeX typesetting language in its documents. AUX files contain information used for cross-referencing, and is also used to transport information from one compiler run to the next.
- Some of the compilers are pdftex, Xelatex, Lualatex etc.
- A log file is usually a flat text file that contains a list of events that happend when a program was running, with one event on each line. Often times errors are recorded in log files.
- .pdf: The common output format for your document. Created by pdflatex/ xelatex

Happy Texing :)

4.6 Introduction to Github

Figure 4.7: Github Logo

GitHub is a Git repository web-based hosting service which offers all of the functionality of Git as well as adding many of its own features. Unlike Git which is strictly a command-line tool, Github provides a web-based graphical interface and desktop as well as mobile integration. It also provides access control and several collaboration features such as wikis, task management, and bug tracking and feature requests for every project.

GitHub offers both paid plans for private repo handle everything from small to very large projects with speed and efficiency. ositories, and free accounts, which are usually used to host open source software projects. As of 2014, Github reports having over 3.4 million users, making it the largest code host in the world.

GitHub has become such a staple amongst the open-source development community that many developers have begun considering it a replacement for a conventional resume and some employers require applications to provide a link to and have an active contributing GitHub account in order to qualify for a job.

The Git feature that really makes it stand apart from nearly every other Source Code Management (SCM) out there is its branching model.

Git allows and encourages you to have multiple local branches that can be entirely independent of each other. The creation, merging, and deletion of those lines of development takes seconds.

This means that you can do things like:

- Frictionless Context Switching.
  Create a branch to try out an idea, commit a few times, switch back to where you branched from, apply a patch, switch back to where you are experimenting, and merge it in.
- Role-Based Code lines.
  Have a branch that always contains only what goes to production, another that you merge work into for testing, and several smaller ones for day to day work.
- Feature Based Work flow.
  Create new branches for each new feature you're working on so you can seamlessly switch

back and forth between them, then delete each branch when that feature gets merged into your main line.

- Disposable Experimentation.
  Create a branch to experiment in, realize it's not going to work, and just delete it - abandoning the workwith nobody else ever seeing it (even if you've pushed other branches in the meantime).

Notably, when you push to a remote repository, you do not have to push all of your branches. You can choose to share just one of your branches, a few of them, or all of them. This tends to free people to try new ideas without worrying about having to plan how and when they are going to merge it in or share it with others.

There are ways to accomplish some of this with other systems, but the work involved is much more difficult and error-prone. Git makes this process incredibly easy and it changes the way most developers work when they learn it.

4.6.1 What is Git?

Figure 4.8: Git Logo

Git is a distributed revision control and source code management (SCM) system with an emphasis on speed, data integrity, and support for distributed, non-linear workflows. Git was initially designed and developed by Linus Torvalds for Linux kernel development in 2005, and has since become the most widely adopted version control system for software development.

As with most other distributed revision control systems, and unlike most clientserver systems, every Git working directory is a full-fledged repository with complete history and full version-tracking capabilities, independent of network access or a central server. Like the Linux kernel, Git is free and open source software distributed under the terms of the GNU General Public License version 2 to handle everything from small to very large projects with speed and efficiency.

Git is easy to learn and has a tiny footprint with lightning fast performance. It outclasses SCM tools like Subversion, CVS, Perforce, and ClearCase with features like cheap local branching, convenient staging areas, and multiple workflows.

4.6.2 Installation of Git

Installation of git is a very easy process. The current git version is: 2.0.4. Type the commands in the terminal:

$ sudo apt-get update

$ sudo apt-get install git

This will install the git on your pc or laptop.

4.6.3 Various Git Commands

Git is the open source distributed version control system that facilitates GitHub activities on your laptop or desktop. The commonly used Git command line instructions are:-

4.6.3.1 Create Repositories

Start a new repository or obtain from an exiting URL

$ git init [project-name]
Creates a new local repository with the specified name

$ git clone [url]
Downloads a project and its entire version history

4.6.3.2 Make Changes

Review edits and craft a commit transaction

$ git status
Lists all new or modified files to be committed.

$ git add [file]
Snapshots the file in preparation for versioning.

$ git commit -m "[descriptive message "]
Records file snapshots permanently in version history.

4.6.3.3 Group Changes

Name a series of commits and combine completed efforts

$ git branch
Lists all local branches in the current repository.

$ git branch [branch-name]
Creates a new branch.

$ git checkout [branch-name]
Switches to the specified branch and updates the working directory.

$ git branch -d [branch-name]
Deletes the specified branch.

4.6.3.4 Synchronize Changes

Register a repository bookmark and exchange version history.

$ git push [alias [branch]]
Uploads all local branch commits to GitHub.

$ git pull
Downloads bookmark history and incorporates changes.

CHAPTER 5

EXPERIMENTAL RESULTS AND COMPARISON

5.1 Experimental Results

We have written some scripts for some basic graph plotting and some analysis.
Below is one of the experimental result.

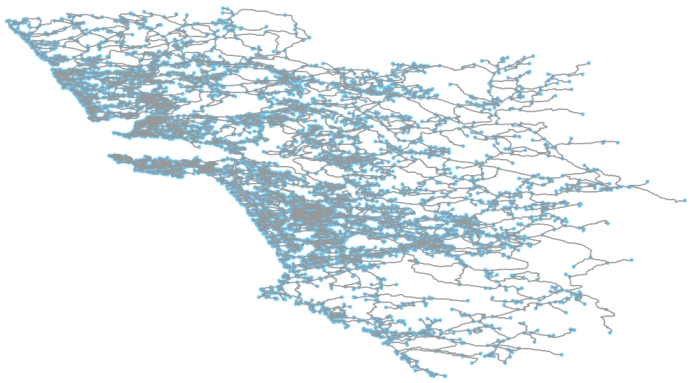

Figure 5.1: Goa Road Network Plotted

You may refer to my blogs also for detailed information.
Here is the url:
http://geoffboeing.com/

Road network of Punjab looks like this:

Figure 5.2: Punjab road network

5.1.0.1 Create Road Networks

Script to create road networks

```
import osmnx as ox

T =”Place osm tag here”

ox.plot_graph(ox.graph_from_place(T))
```

5.1.0.2 Create Road Networks

Script result for finding pagerank using NetworkX:

5.1.0.3 Create Road Networks

Script result for finding pagerank using GraphFrames:

5.2 Conclusions

We concluded some information using processing times and other analysis as follows:

5.2.1 When to use NetworkX

- Unlike many other tools, it is designed to handle data on a scale relevant to modern problems.
- Most of the core algorithms rely on extremely fast legacy code
- Highly flexible graph implementations (a graph/node can be anything!)

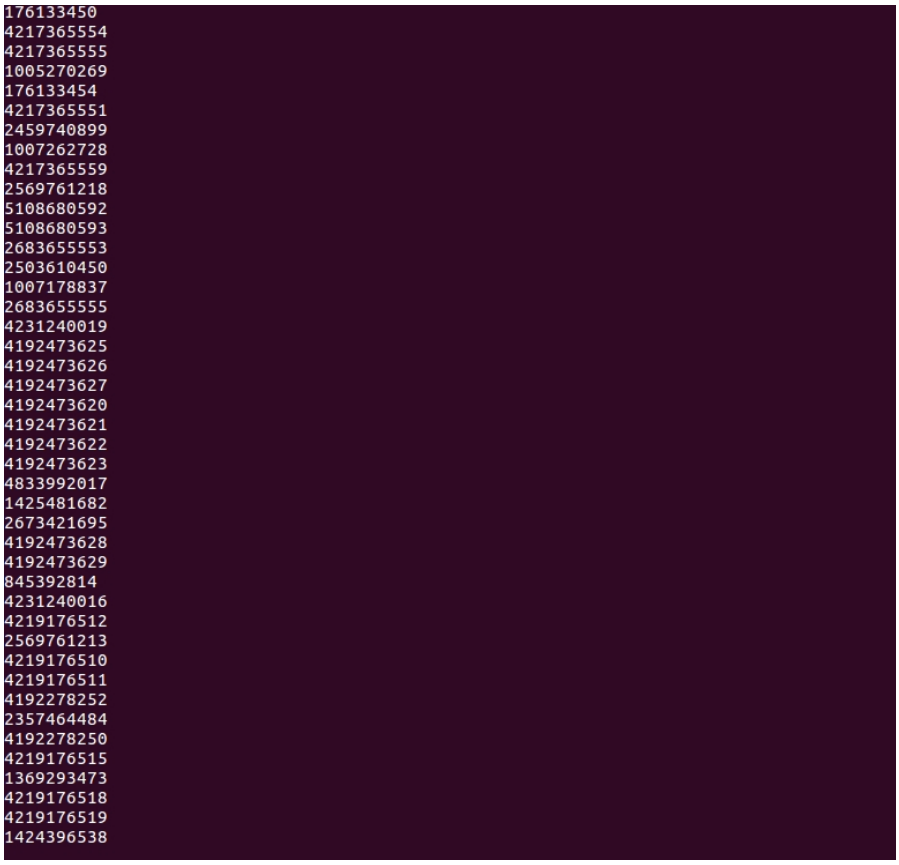

Figure 5.3: Pagerank using NetworkX

- Extensive set of native readable and writable formats
- Takes advantage of Pythons ability to pull data from the Internet or databases

5.2.2 When not to use NetworkX

- Large-scale problems that require faster approaches (i.e. massive networks with 100M/1B edges)
- Better use of memory/threads than Python (large objects, parallel computation)

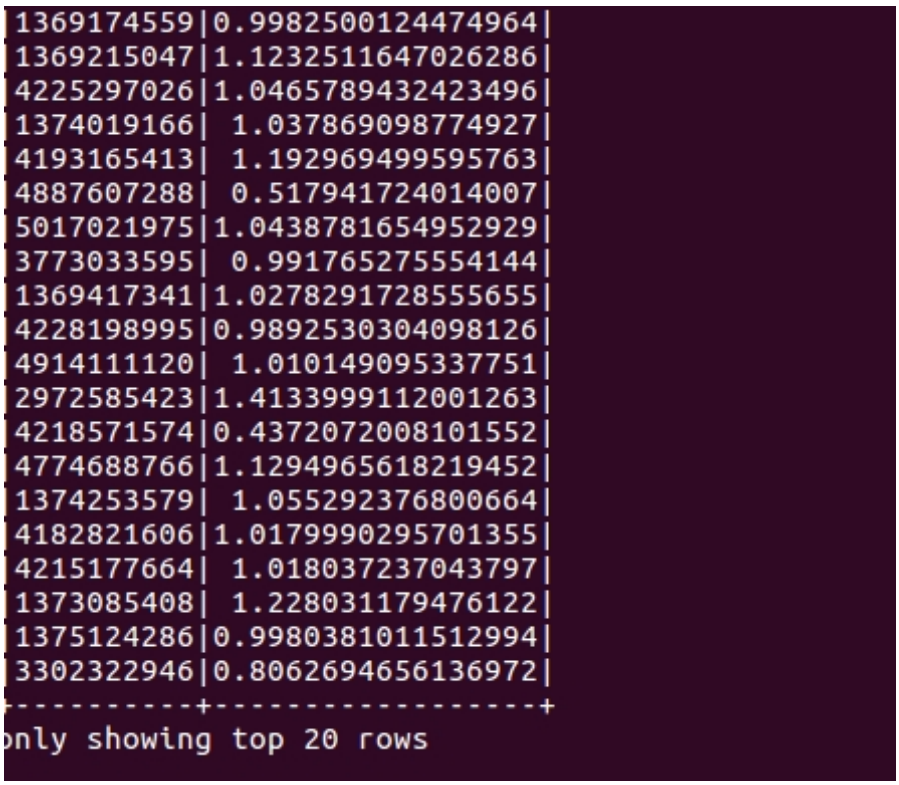

Figure 5.4: Pagerank using GraphFrames

CHAPTER 6

CONCLUSION,SUMMARY AND FUTURE SCOPE

6.1 Future Scope

To analyze OSM data using NetworkX is ideal for small data sets but when data becomes large this approach becomes vague and at that pont we need a new technology whic incorporates Big Data and such technologies are Hadoop, Apache Spark and various others, but Hadoop and Apache Spark are the most popular, Apache spark is 100 times faster than mapreduce pardigm in Hadoop, Graphframes API is used as it supports Python and Larger datasets can be analyzed in future of this project

6.2 Technical and Managerial Lesson Learnt

I learned a lot by doing this project . During this period I got to learn a vast number of technologies. These are listed below :

- **Operating system**: Ubuntu
- **Languages used**: Python
- **Framework**: Reveal.js, Reveal-md
- **Typesetting**: LaTex
- **Other Learnings**: Apache Spark, Markdown

So during this project I learned all the above things. Above all I got to know how Softwares are developed from the scratch. Planning, designing, developing code, working in a team, testing etc. These are all very precious things I got to learn during this period.

BIBLIOGRAPHY

[1] Ntwx, https://github.com/AFCgooner29/Majorproject

[2] LaTeX Beginner's Guide By Stefan Kottwitz

[3] Blog Followed, http://geoffboeing.com/

[4] Our Github Profile, https://github.com/AFCgooner29/ , https://github.com/singh1114/

[5] Online Sources , https://graphframes.github.io/