Dupli Checker

# PLAGIARISM SCAN REPORT



3%
Plagiarised

97%
Unique

| | |
|---|---|
| **Date** | 2022-11-11 |
| **Words** | 754 |
| **Characters** | 7747 |

## Content Checked For Plagiarism

Group Project

19OH01 - Social and Economic Network Analysis

Topic: ENRON E-mail analysis

19Z328-Lavanya Ra
19Z320 - Harshitha P
19Z348 - Sravya V
19Z352 - Sushmitha S
19Z354 - Swetha G N

drawback STATEMENT

The Enron scandal and collapse was one in all the most important company meltdowns in history. In the year 2000, Enron was one in all the most important energy corporations in America. Then, when being outed for fraud, it spiraled downwardintobankruptcywithinayear.WehavetheEnronemaildatabase that contains five hundred thousand emails between one hundred fifty former Enron workers, largely senior executives. It's conjointly the sole giant public info ofrealemails,whichmakesitmorevaluable.
In fact, knowledge scientists are victimization this dataset for education and analysis for years.

DATASET DESCRIPTION

The CALO Project assembled and created this dataset (A psychological feature Assistant that Learnsand Organizes). it's info organized into folders from roughly one hundred fifty users, most of whom area unit high managers at Enron. The corpus has roughly zero.5 million messagesinit.TheFederalEnergy restrictive Commission at the start created this info on the market to the general public and place it on-line whereas conducting its inquiry. Thecollectionconsistsof150individuals'517,431mails,primarily prime management from the Enron Corporation. Despite the dimensions of the gathering, there are

oftentimes few thematic folders sure users. we tend to disregard everyone's inboxes and solely review emails that have already been sent for our functions.
By victimization this strategy, we will stop unintentionallyanalyzingthereceivedemailsthatcontain spam. Sentiment analysis and topic modeling with Latent Dirichlet Allocation (LDA) were the 2 main analytical techniques used.

the essential definition of a social contact between 2 workers could be a pre-defined threshold variety of exchange of emails.
we tend to solely evaluated bifacial links, which means that we tend to take into account a contact if each parties have sent emails to every alternative. This ensures that info was changed between the 2 Enron workers which they were engaged in some reasonably communication. To manifest the flow of knowledge in a corporation, we tend to conjointly thought of every employee's position within the structure hierarchy.

Link to transfer dataset- https://www.cs.cmu.edu/~./enron/

TOOLS USED

- Google Colab
  - Google ColabisafreeJupyterNotebookenvironmentthatrunsentirelyinGoogle Cloud. And utilizes the virtual GPU.
- NetworkX
  - NetworkX could be a python graph library that is employed to construct, manipulate and analyze the structure, options and metrics of the input graph.
- Pandas
  - **The Pandas provides some sets of powerful tools like DataFrame and Series that** primarily used for analyzing the information
- Matplotlib
  - This library is typically paired in conjunction with Numpy to permit users to use python to plot numerous visualizations.

CHALLENGES Janus-faced.

- visualisation of a large dataset. Since most of the nodes within the dataset area unit toorelativeto each other, it's tough to look at them as separate nodes on the screen.

- knowledge gathering techniques and quantifiability problems.

- Missing knowledge verification and proper debugging of errors.

- Somemoduleswereoutdated,sotofindamodulethathasproperdocumentationresearch had to be done.

CONTRIBUTION OFTEAM MEMBERS

Harshitha P - 19Z320 Worked on implementing Page Rank algorithmic program and worked on shrewd basic metrics of the network

Lavanya Ra - 19Z328 Worked on Implementing the HITS Algorithm, Worked on findingouttheCentral and Peripheral nodes

Sravya V - 19Z348 Worked on implementing community detection usingGirvan–NewmanAlgorithm, Displaying the vital nodes supported centrality

Sushmitha S - 19Z352 Worked on implementing community detection usingGirvan–NewmanAlgorithm, property and disconnectivity of Network

Swetha G N - 19Z354 Worked on implementing community detection usingGirvan–NewmanAlgorithm, checking whether or not the network is powerfully connected or bipartite graph

ANNEXURE I: CODE

Import necessary libraries

Import dataset

Basic metrics and finding if its a connected graph and if it's a bipartite graph

Central and boundary nodes

Disconnectivity of graph

agglomeration constant

Degree spatial relation

Betweeness spatial relation

Closeness spatial relation

Page Rank

HITS algorithmic program

Girvan-Newman algorithmic program

distinctive optimum Community split

Markov agglomeration

Label Propagation algorithmic program

Clauset-Newman-Moore algorithmic program

ANNEXURE II: SNAPSHOTS OFTHE OUTPUT
Dataset snapshot:

Network Visualization:

Basic metrics and finding if its a connected graph and if it's a bipartite graph

Central node within the network

boundary nodes within the network

Disconnectivity of graph

Degree bar chart of Network

Degree spatial relation

Closeness spatial relation

Betweeness spatial relation

Page Rank

HITS algorithmic program

Girvan Newman algorithmic program to separate into three communities

distinctive optimum community split