

PSG COLLEGE OF TECHNOLOGY, COIMBATORE 641004

Department of Computer Science & Engineering



**19ZO02 – SOCIAL AND ECONOMIC
NETWORK ANALYSIS**

STAKEHOLDERS COLLABORATION SYSTEM

By

19Z312 - Dharma Dhurai V

19Z326 - Krishna Teja B

19Z338 - Pradeep D

19Z362 - Vinoth Subbiah PL

20Z462 - Mukesh S

BE CSE G2 (2019 – 2023)

PROBLEM STATEMENT

The challenge of handling the new project is to get familiar with the latest technology which is essential for the project. To tackle the project in an efficient and optimal manner we need some technical support. Our proposed system tackles this problem by suggesting a user for collaboration, the topics the user might be interested to learn, and the trending topics using a bipartite graph. The bipartite graph contains two types of nodes: topics and projects. The project has attributes that are project name, user, stars, and forks. These attributes are used for further analysis.

DATASET DESCRIPTION

The dataset is generated by combining multiple CSV files into a single CSV file. After pre-processing the dataset contains seven fields. The dataset used consists of approximately 28280 entries. The Topic field of a particular project consists of the different topics which are all used or involved in the project. A bipartite graph is generated by creating links between the respective Topic nodes and the project nodes. All the topics in the dataset are taken as one partite and the projects are taken as one partite. Based on which the topics a particular project involves the edges are formed. Using Networkx the projection of the graph is made.

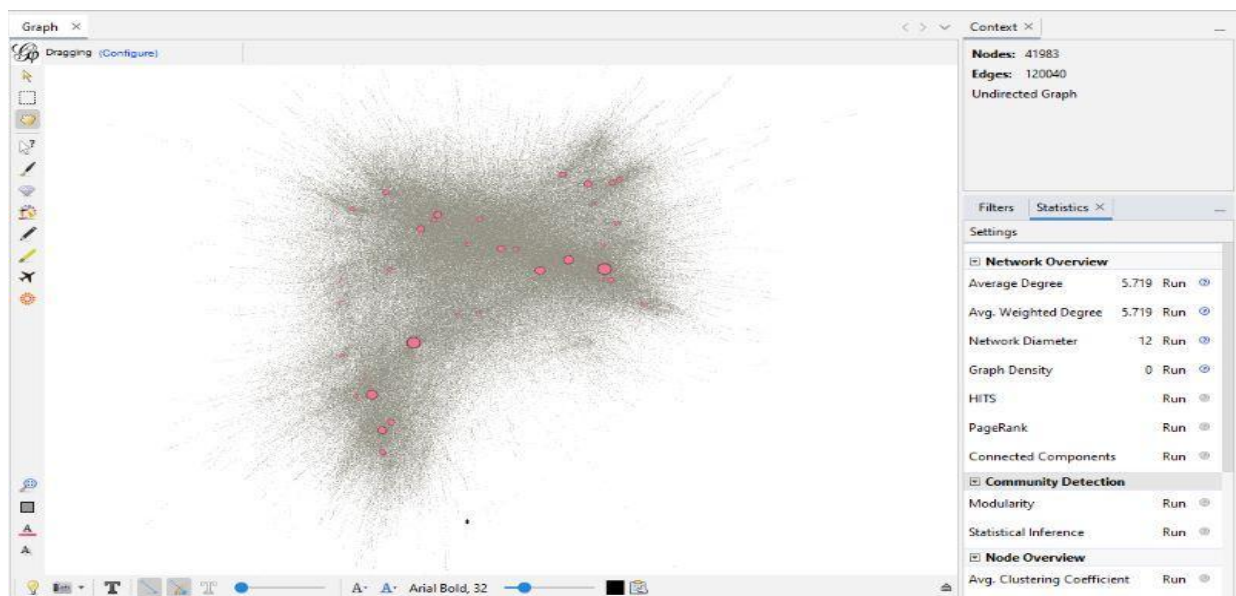


Figure 1 : Network Statistics of the collaboration system.

TOOLS USED

Gephi

Gephi is a visualization application developed in the Java language. It is mainly used for visualizing, manipulating, and exploring networks and graphs from raw edge and node graph data. It is a free and open-source application. It is built on the top of the NetBeans Platform and uses OpenGL for its visualization engine. It runs on Windows, Mac OS X, and Linux. It is an excellent tool for data analysts and data science enthusiasts to explore and understand graphs. It is similar to Photoshop but deals with graph data. The user interacts with the representation and manipulates the structures, shapes, and colors to reveal hidden patterns. The primary goal is to enable the user to make a hypothesis, discover hidden patterns, and isolate structure singularities and defects during data sourcing.

NetworkX

NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and function of complex networks. It is used to study large complex networks represented in the form of graphs with nodes and edges. Using Networkx we can load and store complex networks. We can generate many types of random and classic networks, analyze network structure, build network models, design new network algorithms, and draw networks. Networkx provides classes for graphs that allow multiple edges between any pair of nodes. The Multigraph and MultiDiGraph classes allow you to add the same edge twice, possibly with different edge data. This can be powerful for some applications, but many algorithms are not well-defined on such graphs.

CHALLENGES FACED

- The dataset is gathered in JSON format and the dataset is divided into multiple JSON files we face difficulty in the conversion of the JSON to CSV and while combining the dataset we face difficulty to maintain the overall dataset consistency.
- The dataset consists of different attributes like the user, topic, project, language, etc. it can't be loaded into gephi directly for the visualization of the bipartite graph. So, the dataset attributes are split into topic and project and all other attributes are set as sub-attribute of the project.
- Visualization of the dataset remained a daunting process due to its huge size. Thus, the evaluation of various metrics such as the average path length remained computationally infeasible. Therefore, it required fetching a sub-portion of the dataset without compromising its real-world characteristics such as the degree distribution as per the power law.

CONTRIBUTION OF TEAM MEMBERS

Roll No.	Name	Contribution
19Z312	Dharma Dhurai V	Analysis and generation of user-recommendation system
19Z326	Krishna Teja B	Analysis and generation of the generic filter system, dataset visualization using gephi
19Z338	Pradeep D	Report & Data set collection, Pre-processing and Performance Analysis
19Z362	Vinoth Subbiah PL	Topics recommendation for users based on similar projects done by other users
20Z462	Mukesh S	Report & Data set collection, Pre-processing and Performance Analysis

ANNEXURE I: CODE

GitHub Link:

https://github.com/Krishna-Teja732/stakeholder_collaboration_system.git

ANNEXURE II: SNAPSHOTS OF OUTPUT

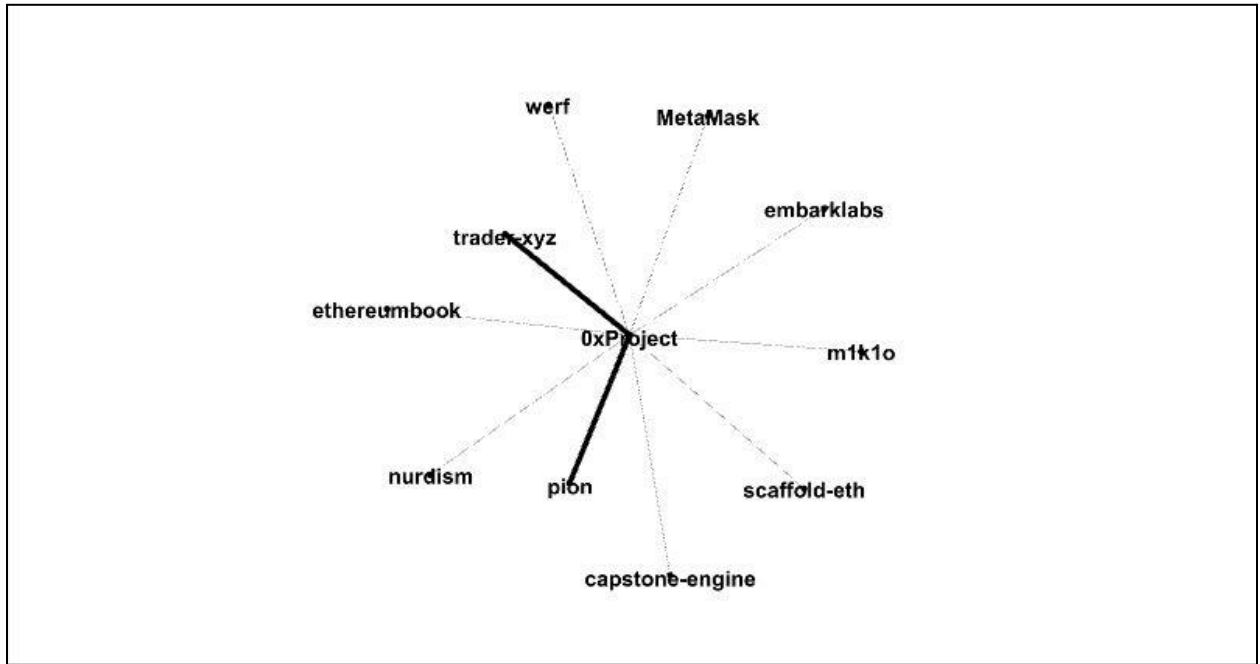


Figure 2: User - user recommendation

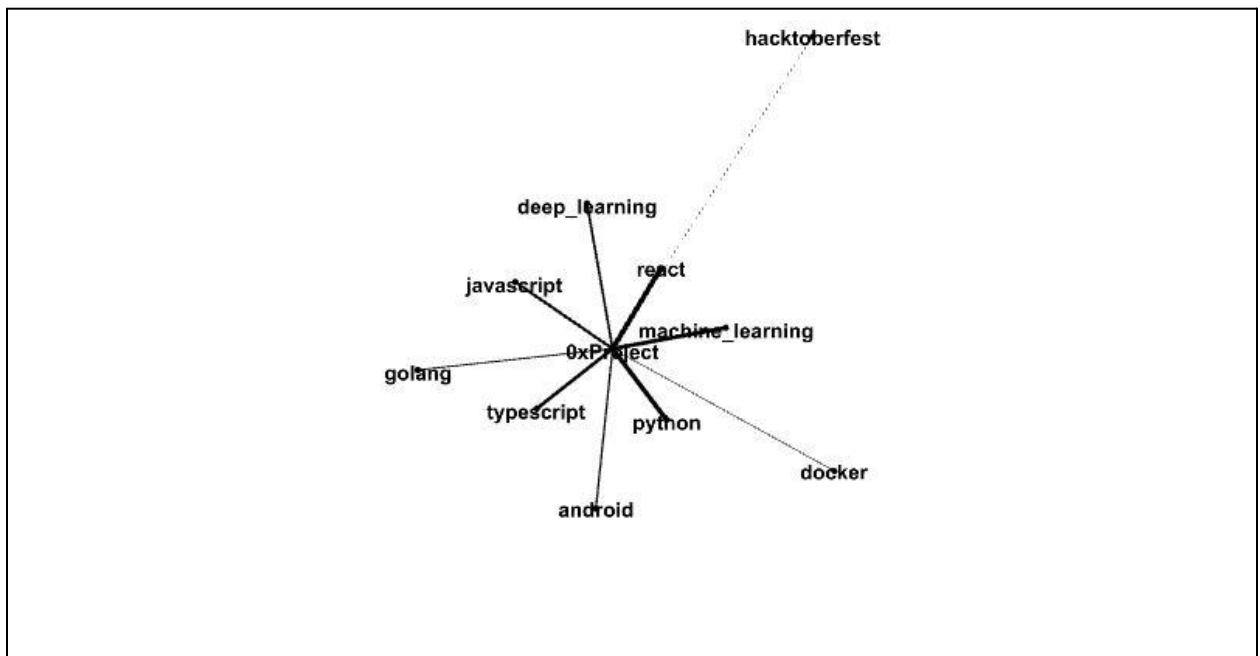


Figure 3: User - topic recommendation

(Thickness of the link indicates the weight of the edges)

REFERENCES

- Dataset: <https://www.kaggle.com/datasets/anshulmehtakaggl/top-1000-github-repositories-for-multiple-domains>
- <https://gephi.org/users/>
- <https://networkx.org/>
- <https://www.datacamp.com/tutorial/networkx-python-graph-tutorial>
- https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf
- <https://github.com/coyotebush/github-network-analysis>
- <https://www.linkedin.com/pulse/github-network-analysis-graph-algorithms-anicet-houkpe/>
- <https://github.com/Dhanya-Abhirami/Social-Network-Analysis-of-Github-Users>
- <https://medium.com/swlh/visualizing-databases-using-gephi-591c9530c981>
- <https://networkx.org/documentation/stable/reference/algorithms/bipartite.html>