JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

Minor Project

Semester V

TOPIC

Community Detection

Team members

Aranya Maji – 21103005

Gargi Jain – 21103018

Ishita Sethi – 21103006

Faculty Supervisor

Dr. Dhanalekshmi G

Motivation behind the project

1. Understanding Network Structure:
    a. Motivation: Investigating the structure of complex networks to gain insights into the relationships and interactions between entities.
    b. Example: Analyzing social networks to understand how individuals are connected and form communities.
2. Identifying Functional Groups:
    a. Motivation: Uncovering groups of nodes with similar functions or roles within a network.
    b. Example: Identifying functional modules in a biological network to understand the roles of different proteins in cellular processes.
3. Improving System Efficiency:
    a. Motivation: Enhancing the efficiency of systems by optimizing the organization and communication within communities.
    b. Example: Optimizing communication within an organization by identifying and strengthening communication channels within specific departments.
4. Detecting Anomalies or Outliers:
    a. Motivation: Identifying unusual patterns or outliers within a network that may indicate anomalies or areas of interest.
    b. Example: Detecting fraudulent activities in a financial network by identifying abnormal patterns of transactions.
5. Enhancing Recommendation Systems:
    a. Motivation: Improving recommendation algorithms by understanding the preferences and connections between users.
    b. Example: Enhancing a movie recommendation system by identifying communities of users with similar movie preferences.
6. Facilitating Targeted Marketing:
    a. Motivation: Targeting specific groups of individuals for marketing campaigns based on their affiliations and interactions.
    b. Example: Customizing marketing strategies by identifying communities of interest within a customer base.
7. Studying Social Dynamics:
    a. Motivation: Investigating the dynamics of social interactions and how communities evolve over time.
    b. Example: Studying how online communities form, grow, and change in response to user interactions.
8. Network Visualization and Interpretation:
    a. Motivation: Enhancing the visual representation of complex networks to aid in interpretation and decision-making.
    b. Example: Creating visualizations of co-authorship networks in academia to understand collaboration patterns.

Details of the project

The programming language used for this project is Python on IDLE(Python 3.11 64-bit) for the Backend functioning of the project and HTML5, CSS and JS for the Frontend Development of the project on Visual Studio Code.

The project allows the user to analyze a network graph csv dataset that is in the required format and provide output graph based on the type of graph and the desired algorithm selected by the user.

Literature on the topic

1. Sathiyakumari, K. and Vijaya, M.S. (2016) Community detection based on Girvan Newman algorithm and link analysis ..., ResearchGate. Available at: https://www.researchgate.net/publication/310736499_Community_Detection_Based_on_Girvan_Newman_Algorithm_and_Link_Analysis_of_Social_Media (Accessed: 29 November 2023).
2. Meo, P.D. et al. (2011) Generalized Louvain method for community detection in large networks, ResearchGate. Available at: https://www.researchgate.net/publication/51929918_Generalized_Louvain_Method_for_Community_Detection_in_Large_Networks (Accessed: 29 November 2023).
3. Cordasco, G. and Gargano, L. (2012) Label propagation algorithm: A semi-synchronous approach - ResearchGate, ResearchGate. Available at: https://www.researchgate.net/publication/264817879_Label_propagation_algorithm_A_semi-synchronous_approach (Accessed: 29 November 2023).
4. Zhang, X. et al. (2009) Modularity optimization in community detection of complex networks, ResearchGate. Available at: https://www.researchgate.net/publication/223956042_Modularity_optimization_in_community_detection_of_complex_networks (Accessed: 29 November 2023).
5. Chejara, P. and Godfrey, W.W. (2017) Comparative analysis of community detection algorithms | IEEE ..., IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/8340627/ (Accessed: 29 November 2023).
6. Mohammad, M. (2018) 'community detection in social networks through Girvan Newman algorithm', Medium. Available at: https://medium.com/@96mehakmohammad/community-detection-in-social-networks-through-girvan-newman-algorithm-78f303912907 (Accessed: 29 November 2023).
7. Yan, M. and Guoqiang, C. (2022) Label propagation community detection algorithm based on density peak optimization, Wireless Communications and Mobile Computing. Available at: https://www.hindawi.com/journals/wcmc/2022/6523363/ (Accessed: 29 November 2023).
8. Zhang, J. et al. (2021) An improved Louvain algorithm for community detection, Mathematical Problems in Engineering. Available at: https://www.hindawi.com/journals/mpe/2021/1485592/ (Accessed: 29 November 2023).

Division of work

1. Aranya Maji – Frontend Development and Research & Information gathering.
2. Gargi Jain – Research and implementation of Louvain's Algorithm and Girvan Newman Algorithm.
3. Ishita Sethi – Research and implementation of Label Propagation Algorithm and Modularity Optimization Algorithm.

Description of the project

Community detection is like viewing a city from above, noticing neighborhoods within the bustling landscape. It's about finding clusters in networks where nodes share stronger connections within their group than outside it. Algorithms act as detectives, uncovering these clusters, revealing how information flows and relationships form. Just as a city evolves, so do these communities, adapting to new connections or changes. Ultimately, it's about understanding the distinct groups that shape the larger network's complexity and dynamics.

The primary goal of this project is to perform, develop an app and provide a research platform for users to work and try different Community Detection Algorithms. Users get to enter their network graph csv dataset that satisfies the format of the project and perform community detection based on what is their network is based on and the suitable algorithms for the same. It allows the users to do a side-by-side comparison of same dataset with four different community detection algorithms applied on it, i.e.:-

1. Louvain's Algorithm
2. Girvan Newman Algorithm
3. Modularity Optimization Algorithm
4. Label Propagation Algorithm

Proposed Methodology

1. The features and algorithms to be included in the project were decided.
2. Based on the features and algorithms, the most suitable programming language was chosen, i.e., Python, as it provided a number of in-built libraries that made working with community detection much easier and made Python a primary choice for Machine Learning developers.
3. Individual program files were made for each of the algorithms and decided upon output graph visualization finally deciding upon Plotly.
4. Upon successful execution of each one of them, each individual algorithms were combined into a single Python program and further minor features were added.
5. Decided upon Frontend Development method where web development was chosen as it provides a much versatile and greater provision for eye catching UI/UX development.
6. Decided and researched on ways to integrate the Python Backend to the HTML Frontend, and finalized on Flask.
7. Created a makeshift Web Design for verification of successful integration of Frontend and Backend.
8. Created a full-fledged Web Design on Figma and used interactive background from CodePen, and implemented it along with the integration features completing Backend, Frontend and their integration with successful execution with multiple permutations of entries and option selections.

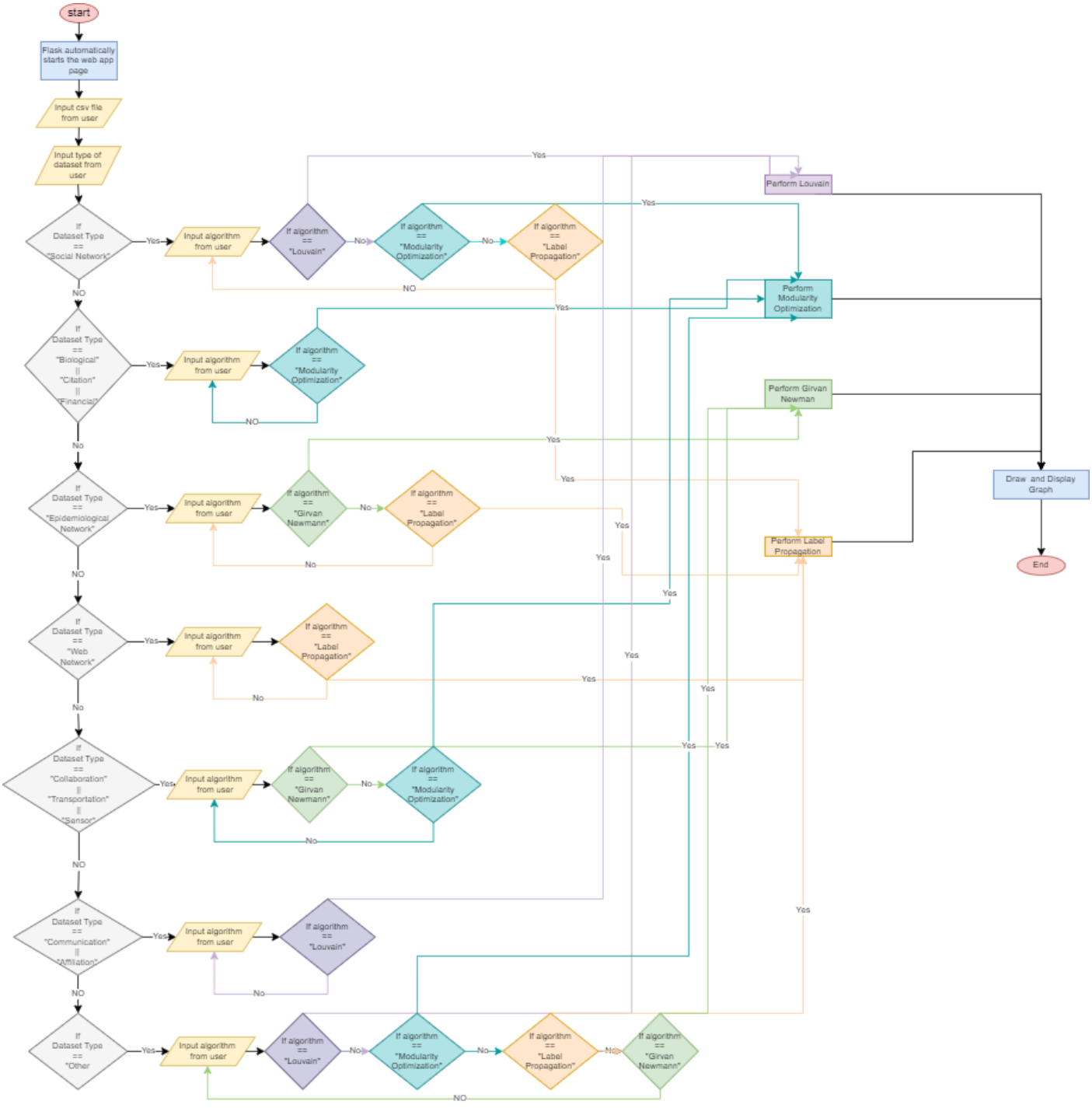
Type of project

Development cum Research Project.

Results

The project is completed with fully functioning Backend integrated with user-friendly Frontend, allowing users to input csv datasets of given format and get network graph visualization with the desired community detection algorithm applied to it.

Overall design of the project



Conclusion

Upon reading all the research papers and implementing the same, we got to know that algorithms like Louvain, Modularity Optimization work well for practically large scale networks as they are fast and algorithms like Girvan Newman and Label Propagation work well with networks having clear community structures.