B.Tech 2020-24 CSE- Project Phase 1

Proposal

I. GROUP NO : DB4

PROJECT TITLE : Unraveling Network Structures: A Leader-Centric Approach to Community Detection in Attributed Networks.

TEAM MEMBERS :

ROLL NO.	NAME
AM.EN.U4CSE20339	Krishnpriya Dinesan
AM.EN.U4CSE20366	Sreesankar S
AM.EN.U4CSE20321	Devesh Kumar V V
AM.EN.U4CSE20152	Pranav B Nair

II. Abstract

The problem to be addressed in this project is the lack of an effective community detection algorithm that can take advantage of both topological and attribute information in attributed networks. The motivation for this project is that community detection is an important task in many domains, such as social network analysis, recommender systems, and fraud detection. However, existing community detection algorithms that only use topological information or attribute information are often not effective.

The persisting challenges in addressing this problem include:

- How to effectively combine topological and attribute information to improve the accuracy of community detection.
- How to deal with the challenges of scalability and efficiency in large-scale attributed networks.

III. Background Study

The following are five papers that align with the project:

Title & year	Problem	Contributions	Limitations	Open problems/Future work
Leader-aware Community Detection in Complex Networks [2019]	The problem addressed in this research paper is community detection in complex networks, where the goal is to identify groups of nodes with strong intra-group connections and weak inter-group connections.	The paper proposes a leader-aware community detection algorithm that incorporates the concept of leadership to improve the accuracy of community detection. The algorithm assigns leadership scores to nodes based on their structural characteristics and utilizes these scores to guide the community detection process.	One limitation of the proposed algorithm is that it assumes a static network, where the connections between nodes remain constant throughout the analysis. Additionally, the algorithm may not perform optimally on very large networks with millions of nodes and edges.	The research paper identifies a few open problems and areas for future work. Some of these include investigating the algorithm's performance on dynamic networks, exploring the impact of different leadership score definitions, and enhancing the algorithm's scalability for large-scale networks.
Evaluation of Community Detection Methods. [2019]	The problem addressed in the paper is the evaluation of community	The paper provides a comprehensive evaluation and comparison of	The paper acknowledges that the evaluation is based on a specific set of	The paper suggests several areas for future research, such as developing more

	detection methods.	various community detection methods. It examines their performance on different types of networks and datasets. The authors analyze the strengths and weaknesses of the methods and provide insights into their applicability in real-world scenarios.	algorithms and datasets, so the findings may not generalize to all community detection methods or network structures. The authors also point out the limitations of current evaluation metrics and discuss the challenges in comparing different methods accurately.	robust evaluation metrics and benchmark datasets to enable fair comparisons between methods. It also highlights the need for studying the performance of community detection methods in dynamic networks and exploring the potential applications of community detection in various fields.
Leader Similarity Based Community Detection Approach for Social Networks [2020]	The paper addresses the challenge of simultaneous community detection and leader selection in social networks without prior knowledge of community sizes and numbers.	The paper proposes the "Leader Similarity Based Community Detection (LSBCD)" algorithm, utilizing similarity measures to identify influential leaders and form communities around them.	The LSBCD algorithm's focus is on non-overlapping communities, which means the algorithm may not be suitable for detecting overlapping communities in complex networks. Another focus is on the applicability to extremely large	Extending the LSBCD algorithm to handle overlapping communities in social networks, which would enhance its applicability to real-world scenarios. Further exploring the performance of the LSBCD algorithm on other types of social networks, such as

			and dynamic networks, such as those found in real-time social media platforms, is not explored as much.	online social media platforms like Facebook and Twitter, which are characterized by their dynamic nature and large user base. Investigating the LSBCD algorithm's scalability and efficiency on extremely large networks to determine its applicability to big data scenarios.
Deep Learning Techniques for Community Detection in Social Networks. [2020]	The problem addressed in the research paper is likely how to leverage deep learning techniques to improve the accuracy and efficiency of community detection in social networks.	The paper proposes novel deep learning models tailored for social network community detection, integrating network information, enhancing scalability, robustness to noisy data, transfer learning, interpretability, and demonstrating performance	Deep learning techniques for community detection in social networks face challenges: large, labeled data requirements, high computation complexity, lack of interpretability, overfitting risk, limited generalization, hyperparameter selection difficulty, explain ability issues,	Future work in deep learning for community detection includes improving semi-supervised and unsupervised methods, handling dynamic and heterogeneous networks, addressing fairness and bias, enhancing scalability and efficiency, integrating network

		improvements over traditional methods.	imbalanced community sizes, and potential privacy concerns.	embeddings, advancing interpretability, investigating adversarial robustness, exploring multi-view approaches, and applying to real-world scenarios.
Leader-Based Community Detection Algorithm in Attributed Networks [2021]	The problem statement mentioned in the introduction of the research paper is the need for an improved community detection method that incorporates both topological and attribute information to maintain the integrity of information in complex networks.	The paper introduces TALB, a leader-based method that integrates topological and attribute information to establish node dependencies. By combining attribute similarity matrices with network topology, a dependency tree is formed, enabling improved community division. Experiments on synthetic and real networks demonstrate the method's superiority over	TALB method faces efficiency and universality challenges in large, dense networks. Incorporating topological and attribute information may increase computational overhead, while effectiveness depends on specific network characteristics. Addressing this requires efficient algorithms and extensive evaluations for real-world applicability.	Develop more efficient algorithms or optimization techniques that can handle large and dense complex networks without sacrificing accuracy. This could involve parallelization, distributed computing, or sampling approaches to improve scalability. Also investigate the robustness of the TALB method under different perturbations, such as node removal, edge

		existing algorithms.		addition, or changes in attribute data. Understanding how the method performs under varying conditions can provide insights into its reliability and generalization.
Community Detection in Attributed Networks Using Graph Wavelets. [2022]	The paper addresses the graph signal processing-based approach to community detection in attributed networks.	The proposed algorithm uses spectral graph wavelets to filter the attributes and constructs a new network from the graph filtered attributes across different scales. In this manner, both the graph connectivity information and the node attributes are taken into account in the community detection task. The proposed method is evaluated on multiple attributed social networks and is shown to	Community detection in attributed networks using graph wavelets has limitations including computational complexity, sensitivity to graph structure, and challenges with continuous attributes and attribute dependencies. Interpreting wavelet coefficients and evaluating results can be difficult. Improving scalability, accuracy, and interpretability	The proposed method detects communities taking into account both the node attributes and the topology of the network. Two classes of graph filters based on graph wavelet transform were implemented, a band-pass wavelet filter and a low pass scaling filter. Experiments on several real-world datasets with binary and numerical attributes were conducted. The experiments show that using the

		perform well on networks with both binary and numerical attributes.	remains an active research area.	attributes in addition to the topology yields more accurate communities. Moreover, results show that the low pass scaling filter performs better than the band-pass wavelet filter, but both of them outperform the conventional clustering algorithms. Future work will consider other kernel functions and the extension of this framework to detect multi-scale community structure.

IV. Challenges

Based on the background study, the following are these are the challenges that still exist in leader-based community detection algorithms in attributed networks:

- How to effectively select leaders in the network.
- How to deal with the challenges of scalability and efficiency in large-scale networks.
- How to incorporate topological information into the algorithm.

V. Deliverables of Phase I

The deliverables of Phase I of the project are: • A literature survey that identifies six most suitable base papers that align with the project. • A high-level design of the project that clearly exhibits inputs/outputs as well as the modules that are planned to be implemented as part of the Project Phase - I. • A prototype implementation of the algorithm.

VI. Assumptions/Declarations:

The following assumptions are made for the project: • The network is undirected and unweighted. • The attribute information is categorical. • The dataset is available in a public repository.

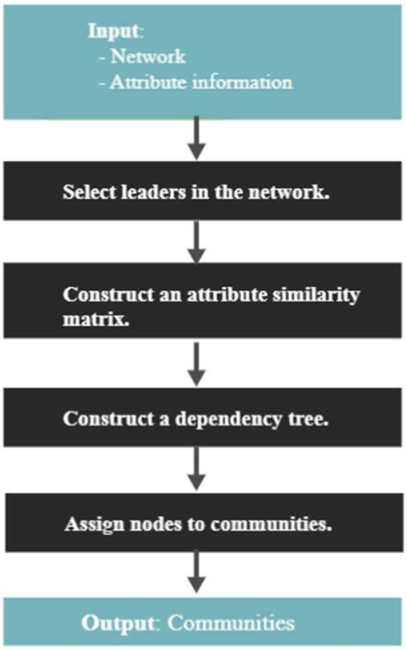
VII. Tools to be used.

Software/Hardware Tools	Specifications
The following tools will be used	**Python :** High-level, versatile programming language with simple syntax, ideal for web development, data analysis, and automation.

for the project: ● Python ● NetworkX ● Scikit-learn	NetworkX : Graph analysis library for Python; used to create, manipulate, and study complex networks. Scikit-learn: Machine learning library in Python with simple and efficient tools.

VIII. High Level Design

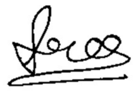
The high-level design of the project is shown in the following block diagram:



The project will be implemented in Python using the NetworkX and Scikit-learn libraries. The dataset will be available in a public repository.

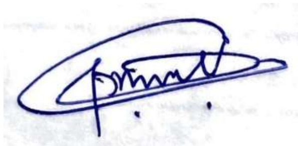
Students’ Name and Signature :

KRISHNPRIYA DINESAN



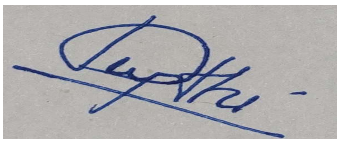
SREESANKAR S

DEVESH KUMAR V V



PRANAV B NAIR

Guide’s Signature



DEEPTHI L R