B.Tech 2020-24 CSE- Project Phase 1

Proposal

I. **Group No:** B5

Project Title.:

Multilevel attribute inference mechanism on attributed social networks

Team members:

Roll No.	**Name**
AM.EN.U4CSE20175	YADDALA PAVAN KUMAR REDDY
AM.EN.U4CSE20148	NUKALA AASHISH GOPAL
AM.EN.U4CSE20164	SIRIGUPPA VISHNU SAI KAUSHAL
AM.EN.U4CSE20147	NETTEM NIKHIL CHOWDARY

II. Abstract

In social networks, users possess hierarchical user attributes that serve as self-descriptive information. Attribute inference plays a crucial role in social networks for user classifications and targeted recommendations. However, current approaches mainly focus on the flat inference problem, leading to significant inconsistencies in multi-level tasks. This project addresses this issue by proposing a solution i.e., Multi-level Inference model, that takes into account the semantic hierarchy of user attributes. By considering the hierarchical structure, this approach aims to improve the accuracy and reliability of attribute inference, enabling more consistent and effective user classifications and targeted recommendations on social platforms.

Attribute inference is relevant because it enables social platforms to improve user experiences, provide personalized services, optimize platform operations, and drive targeted advertising, ultimately enhancing the value and effectiveness of social network platforms.

Motivation:

The motivation behind this project is to address the limitations of existing approaches in attribute inference for user attributes in social networks. Currently, most methods focus on flat inference, disregarding the semantic hierarchy of user attributes. For example, consider a social networking platform where users have attributes related to their hobbies and interests. With single-level inference, the platform may only be able to infer broad categories such as "sports" or "music" based on user behavior. However, it may miss out on specific subcategories like "soccer" or "jazz music" that provide more accurate insights into user preferences. In this case, multi-level inference becomes crucial. It allows the platform to infer attributes at different levels of granularity, considering the hierarchical structure. By capturing both broad and specific attributes, the platform can offer more precise recommendations and personalized experiences to users.

III. Background Study

Title &year	Problem	Contributions	Limitations	Open problems/Future work
MLI: A Multi-level Inference Mechanism for User Attributes in Social Networks Year of Publication: 2022	In recent years, there has been a growing focus on user attribute inference, with particular concern regarding flat inference, where attributes are at the same level. Current approaches mainly focus on the flat inference problem, leading to significant inconsistencies in multi-level tasks. So Inferring multi-level attributes can make the inference results more accurate and	Introduced MLI model for multi-level attribute inference in social networks, utilizing a "Ripple" algorithm and a generalized semantic tree. Experiments validate its superiority on real data sets.	MLI does not infer user attributes accurately all the time but fails in some special cases. Since not all users in social networks satisfy the principle of homogeneity, it also leads to the incorrect inference result by MLI	Future work of this paper includes automating the learning of high-quality semantic trees from data for improved applications. They aim to utilize advanced GNN models to analyze social network structures and use GNNs for inference. Exploring the use of users' textual content to enhance

	analyze users more effectively			accuracy is also a priority.
Inference of personal attributes from tweets using machine learning Year of Publication: 2017	Predicting personal attributes from social data is a prominent research focus, hindered by data availability constraints. In the social media era, people willingly share linguistic expressions, providing valuable information for attribute inference. Computational methods for inferring personal attributes from social data are relatively unexplored, lacking a versatile algorithm. This research aims to explore predicting attributes solely from social media text.	The study explored the impact of word embedding dimension (N), tweet block size (L), and various algorithms on the prediction accuracy of gender, occupation, and age groups. Two preliminary results were reported for predicting personal attributes from text tweets using five different machine learning algorithms.	The machine learning algorithms demonstrated a prediction accuracy of just 60-70% for the three specific personal attributes (gender, occupation, and age groups) of interest.	In future research, the authors had planned to enhance deep learning algorithms using larger datasets to conduct prediction tests for various personal attributes. Once a general framework for attribute inference is established, it can be applied to diverse fields, such as computational social science research and personalized marketing applications.
A Temporal User Attribute-Based Algorithm to Detect Communities in Online Social Networks	This paper addresses several limitations of existing community detection algorithms for online social networks. Specifically, it takes into account the temporal and user attributes of social	1) The study developed a new algorithm that addresses the time-varying nature of OSNs by taking into account some user attributes such as users' weight, the density of	The algorithm that was evaluated using a single dataset from a specific online game, and it is unclear how well it would perform on other types	1. Investigating the performance of the proposed algorithm on other types of online social networks and datasets. 2. Exploring the use of other fitness metrics for

Year of Publication: 2020	networks, which are often overlooked by existing edge-based algorithms. Additionally, it aims to be more human-centric in its approach to community detection, which is not always the case with existing algorithms.	interaction, geo-location, and users' lifespan. 2) The study designed a Maximum Spanning Tree (MST) similar to the study in, as an output of the proposed algorithm to reduce the search complexity such as required to reduce the amount of memory necessary for storing information among the communities.	of online social networks. Additionally, the paper does not provide a detailed analysis of the algorithm's computational complexity or scalability, which could be a potential limitation. Finally, the proposed algorithm may require a significant amount of user attribute data, which may not always be available or easy to obtain.	community detection in online social networks, beyond modularity. 3. Developing more efficient and scalable algorithms for community detection in online social networks. 4. Investigating the impact of different user attributes on community detection performance and exploring new attributes that could be used.
Joint Item Recommendation and Attribute Inference: An Adaptive Graph Convolutional Network Approach	The paper addresses the problem of joint item recommendation and attribute inference in recommender systems. Specifically, given user and item attribute information, the paper aims to recommend items to users and predict missing attribute values for either users	Proposed an Adaptive Graph Convolutional Network approach to joint item recommendation and attribute inference in recommender systems. It can handle both user and item attributes and can predict missing attribute	The proposed approach may not be suitable for all types of recommender systems or datasets. It is always important to evaluate the performance of any proposed approach on different	Exploring the use of additional types of auxiliary data for recommendation, such as images and audio, and investigating the use of more advanced graph neural network architectures. Investigating the use of the

Year Of Publication: 2020	or items. The proposed approach is based on an Adaptive Graph Convolutional Network.	values for either users or items. The paper also presents an extensive experimental evaluation of the approach on several benchmark datasets, demonstrating its effectiveness compared to other state-of-the-art methods.	datasets and compare it with other state-of-the-art methods to better understand its limitations.	proposed approach in more complex scenarios, such as cold-start recommendation and dynamic recommendation.
Active learning for hierarchical multi-label classification Year Of Publication: 2020	This paper focuses on the problem of hierarchical multi-label classification (HMC) and proposes a new active learning algorithm called Hierarchical Query-By-Committee (HQBC) to address this problem. The authors also provide a comprehensive framework for active learning in HMC and evaluate the performance of various algorithms on several benchmark datasets.	1. A comprehensive framework for active learning in hierarchical multi-label classification 2. A new active learning algorithm called Hierarchical Query-By-Committee (HQBC) that is specifically designed for HMC and outperforms existing methods in terms of predictive performance and computational efficiency.	The paper does not provide a detailed analysis of the computational complexity of the proposed algorithm, which could be a limitation in some settings.	1. Incorporating datasets with more complex label hierarchies, such as direct acyclic graphs, and developing AL algorithms that can consider many hierarchical paths from the root node to any class. 2. Extending the proposed framework to address multi-target regression and hierarchical regression problems.

IV. Challenges

1. When the existing single-level methods are directly applied to infer attributes at each level of the hierarchy, three issues commonly arise: conflicts, indeterminacy, and missing values. These problems occur due to the failure to consider the hierarchical relationships between attributes, leading to inconsistencies, uncertainty, and incomplete attribute inference results.
2. Existing methods may not be able to handle the dynamic nature of social networks, where user attributes and social connections can change over time.
3. The limitations of information collection and the presence of noise in social networks result in incomplete attributes. Only a fraction of users possess known attributes, while the majority remains unknown.

V. Deliverables of Phase I

Objectives:

1. Develop a model for inferring hierarchical attributes for unknown users in social networks.
2. Design an algorithm to accelerate the inference and improve the effect.
3. Evaluate the effectiveness of the model on real data sets and demonstrate its superior performance compared to other existing methods.

Outcomes/Deliverables:

1. A model for inferring hierarchical attributes for unknown users in social networks.
2. An algorithm to accelerate the inference and improve the effect.

VI. Assumptions/Declarations:

- One of the main assumptions is that the user attributes to be inferred have a hierarchical structure with multiple levels.

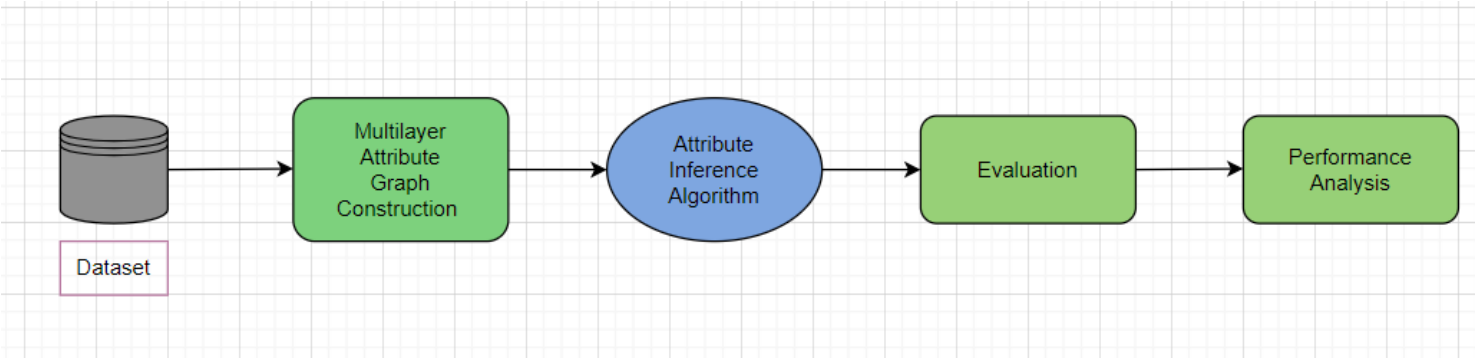
Dataset:

- DBLP (Digital Bibliography & Library Project) dataset. DBLP is a database system of English essays in the computer field, where each user is a vertex and co-authors are connected by an edge and their research fields are used as the attributes to be inferred.

VII. Tools to be used

Software/Hardware Tools	Specifications
Python	Version 3.11

VIII. High Level Design



Students' Name and Signature

YADDALA PAVAN KUMAR REDDY - Y. Pkreddy

NUKALA AASHISH GOPAL - N. Aashish_gopal

SIRIGUPPA VISHNU SAI KAUSHAL - Kaushal

NETTEM NIKHIL CHOWDARY - Nikhil Chowdary

Guide's Signature

(Deepthi L R)

Assistant Professor (Sr. Grd.),

School of Computing, Amritapuri.