Final Year Project Plan

Interaction of LLM and Graph Learning

Zhao Jieyi

zjy800@connect.hku.hk

Supervisor: Prof. Huang Chao

1. Project Background

1.1 Background

Currently, LLMs have demonstrated outstanding potential across various domains, as they have shown significant ability in summarizing and logistics, which are desired in many fields. They can also help understand the textual information and capture contextual relationships without using human labor. The recent LLM research is increasing as the quickly growing computational power of hardware and its usage in other fields are well noticed by researchers.

Graph learning, which has been widely used in many areas, like social network analysis and recommendation systems, can show the connection and interactions between items. The existing graph learning algorithms are well-designed to capture the structural information of the interaction graphs. Nowadays, graph learning algorithms are widely studied not only in research fields but also in industrial and application fields.

Both these fields have shown promising futures, and their integration can also be promising. By combining LLMs with graph learning, the performance on graph-related tasks can be improved, as integration could lead to more contextually rich and structurally aware models and take better advantage of textual data and semantic data, which were previously hard to use in graph learning fields. Equipped with semantic embedding, graph learning can better understand the real-life situation and get better performance.

1.2 Research Gap

Although the field is promising, the existing methods cannot fully utilize the contextual understanding of LLMs and combining with graph learning. The optimal features to extract from textual data that can best complement existing graph learning algorithms are yet to be determined. Besides, the researchers are trying to find a better

structure to combine these algorithms.

Based on the limitations of the current study, the project aims to explore multiple integration methods and find features to extract from the textual data and combine them with the existing graph learning algorithms. The research will also try to develop new models and frameworks to help the LLMs better understand graph structure, so that the integration will be more practical and gain better performance.

1.3 Related Work

Representation Learning with Large Language Models for Recommendation [1]

The research has explored the integration of Large Language Models (LLMs) with recommendation systems to enhance their performance. It utilizes the feedback of user-item interaction and embeds the feedback contents to get the semantic data. RLMRec represents a significant way towards effectively combining the strengths of graph-based methods and LLMs in recommendation systems, showing the potential for improved performance and robustness in handling real-world data challenges.

GraphGPT: Graph Instruction Tuning for Large Language Models [2]

GraphGPT is a framework that aims to align Large Language Models (LLMs) with graph structural knowledge using a graph instruction tuning paradigm. The framework introduces a text-graph grounding component to establish connections between textual information and graph structures. This is important for bridging the gap between natural language understanding and graph representation learning.

2. Project Objective

Explore and develop integration methods for LLMs and graph learning

- Develop new techniques for using LLMs to combine graph node information with textual data
- Create methods to incorporate graph structure into LLM inputs for improved contextual understanding

Identify and extract relevant features from textual data

- Determine which features extracted by LLMs are most relevant and beneficial for graph learning tasks
- Design prompts to get useful responses from LLMs that focus on valuable features.

- Develop methods to efficiently extract and utilize these features in graph learning algorithms.

Enhance existing graph learning algorithms with LLM capabilities

- Incorporate LLM-derived textual features into current graph learning models
- Improve the semantic understanding of graph structures using LLM-based approaches
- Use LLMs for data preprocessing and data augments to make the graph learning models robust.

Develop new models and frameworks.

- Create innovative architectures that seamlessly integrate LLMs and graph learning.
- Design models that can take semantics meanings into account to gain better performance.
- Develop model-free frameworks that allow for flexible integration of different LLMs with various graph learning algorithms.

Evaluate and benchmark the integrated approaches.

- Compare the new integrated approaches with existing baselines and state-of-the-art methods.
- Ablation studies will be conducted to understand the contribution of each component in the integrated models.

Investigate real-world applications

- Apply the developed methods to practical scenarios such as recommendation systems.
- Analyze the effectiveness of the integrated approach in improving performance on these real-world tasks.

3. Project Methodology

The project on the interaction of Large Language Models (LLMs) and Graph Learning will be implemented systematically, combining theoretical research, practical experimentation, and iterative development. Here's an overview of the implementation methodology:

Investigation of previous related research

As LLM has gained significant attention for its potential to improve various fields, numerous works have emerged focusing on LLM and graph learning. To gain a deeper understanding and develop my own contributions, I will examine related works in our Lab and other research groups.

Data investigation

Exploring datasets suitable for graph-based tasks will be a key part of this research. These tasks include social network analysis, recommendation systems, and knowledge graph completion. Many existing datasets used for graph learning are also well-suited for LLM applications. By studying these datasets, The project aims to enhance my understanding of graph learning, LLM, and their interaction.

Integration of LLMs and graph learning

Several methods for integrating LLMs and graph learning will be investigated during this research:

- Using LLMs to enrich graph nodes with textual data, I will explore Graph Augmentation with Natural Language. This approach provides additional context for graph algorithms.
- To improve understanding and generation of text based on graph relationships and entities, I will work on incorporating graph structure into LLMs' input, focusing on Graph Representation in LLMs.
- Valuable mechanisms from LLM can be implemented in graph learning. For instance, the attention mechanism can effectively capture informative parts while reducing noise interference.

Model development

Combining existing LLM models and algorithms with graph learning will be a focus during my research internship. By experimenting with these combinations, I hope to assess their performance. My goal is to develop new models or frameworks that merge graph learning with LLM, aiming to enhance performance in graph learning tasks.

Evaluation

To ensure stability and versatility, the developed models will be evaluated on multiple datasets. Various metrics appropriate to the tasks will be used, and the results will be compared against baseline models that don't use the integrated approach.

Validation

The research validation will involve three main approaches:

- Comparing our methods' performance against established benchmarks in graph learning and LLM applications through benchmarking.
- Applying our methodologies to real-world scenarios, such as social network analysis, misinformation detection, and recommendation systems.
- Conducting detailed ablation studies to understand each component's contribution to the integrated model.

Expected outcomes

Novel methodologies for integrating LLMs with graph learning are expected to result from this research. These should lead to improved performance on graph-related tasks. Additionally, insights into how these two fields can mutually enhance each other should be gained.

4. Project Schedule and Milestone

Schedule	Milestones
Sept. 2024- Oct. 2024	Investigating the previous related research
Nov. 2024	1. Data investigation and preprocessing 2. Algorithm design
Dec. 2024 – Jan. 2025	Model development
Feb. 2025 – Apr. 2025	Experiment
May. 2025	Report summarizing

References:

[1] X. Ren et al., “Representation Learning with Large Language Models for Recommendation,” arXiv.org, Feb. 25, 2024. https://arxiv.org/abs/2310.15950 (accessed Apr. 09, 2024).

[2] J. Tang et al., “GraphGPT: Graph Instruction Tuning for Large Language Models,” arXiv.org, Oct. 19, 2023. https://arxiv.org/abs/2310.13023