UNIVERSITY OF SCIENCE, VIETNAM NATIONAL UNIVERSITY

CSC17103
GRAPH MINING

FACULTY OF INFORMATION TECHNOLOGY
COMPUTER SCIENCE DEPARTMENT

LAB INSTRUCTOR:

FINAL PROJECT

Revision History

Revision	Date	Author(s)	Description
1.0		LN Thanh	Initial handout.
1.1		LN Nam	Vietnamese version.
2.0	May 9, 2024	LN Nam	English version. Modified and added new requirements in all sections.

Contents

1 Learning Objectives

The overarching objective of this project is to cultivate proficiency in research methodologies within the domain of **graph machine learning and network sciences**. Through comprehensive examination of both conventional and contemporary methodologies throughout project execution, participants are anticipated to develop adeptness in scientific inquiry encompassing literature review, report composition, coding demonstrations, and presentation delivery.

2 Notes and Constraints

> ✏️**Note the following:**
> You have to follow these listed things strictly.

🖌️ This project must be done by your team! NOT other team.

🖌️ Work without a report will not be graded.

🖌️ Members who do not contribute to the project will not receive points.

🖌️ Reference sources (if any) need to be fully recorded in the report in the References section. Note that it is necessary to distinguish between referencing and plagiarism.

🖌️ Individuals or groups that commit cheating and dishonesty will receive 0 points in the course.

🖌️ If the size of submission is > 20MB, upload it to an external storage service such as Google Drive or OneDrive, and then submit the link. Last but not least, please keep the link public for at least 2 years.

3 Required Software and Tools

🖳 **Operating system:** 🐧 Linux (it's better than the remain ones), ⊞ Windows 10, **or** 🍎 macOS

</> **IDE:** Vim, Visual Studio Code, etc.

🌐 Web Browser: Google Chrome.

📑 Markdown, LaTeX for writing documentation.

git Distributed version control system.

🐙 Github: a web-based version control repository hosting service.

📖 Trello: a Web-based project management system.

🔌 A storage medium (a USB flash memory or any online free storage service such as GDrive or OneDrive) for storing and backing up your files.

4 Overview the project

This project is seperated into two main parts:

- Task 1 (25%): Studying and demonstrating algorithms, methodologies, and techniques that were presented by the lecturer.
- Task 2 (75%): Researching the recent advances in the graph deep learning domain.

5   Task 1

For this assignment, each team will be randomly assigned three techniques by the teaching assistant. Your team's responsibility will be to thoroughly study each assigned technique and subsequently present your findings in a comprehensive report, complete with illustrative examples to elucidate the workings of each algorithm/technique.

The structure of the report should encompass the following components:

- *Group Information*: Provide pertinent details regarding the members of your team.
- *Current Status for Each Requirement*: Offer an overview of the current progress status pertaining to each assigned requirement.
- *Brief Introduction to Background and Preliminaries*: Furnish a succinct introduction delineating the relevant background information and foundational concepts pertinent to the assigned techniques.
- *Introduction of Assigned Techniques*: Present a detailed exposition of the techniques assigned to your team. It is strongly advised to incorporate illustrations and examples within the report to facilitate comprehension.
- *References*: Include a comprehensive list of references utilized in the compilation of the report.

Adhering to this structured format will ensure clarity and coherence in the presentation of your team's research findings.

6   Task 2

Groups are instructed to meticulously search for and select a scholarly article within the domain of graphs, graph data mining, or machine learning with graph data. The purpose of this selection is to facilitate in-depth research, experimentation, and evaluation.

You can refer the papers that are listed in these link below:

- Link 1: Graph Deep learning Literature on Top tier conferences
- Link 2: Papers on Graph Analytics

The criteria for paper selection are as follows:

- The chosen article must be distinct from those selected by other groups, ensuring diversity of topics. To ensure this, teams are required to consult the Google Sheet on Moodle.
- The selected article should not be a survey paper, but rather an original research contribution.
- Preference is given to articles published recently, with a focus on the period from 2020 to the present, to ensure relevance and currency of information.
- In cases where the publication date predates 2017, the chosen article must be a research paper specifically addressing algorithms. Additionally, teams are expected to highlight any advancements or updates made in recent years.
- Groups have the option to select an article from the provided link or propose their own. However, the final decision regarding the group's choice will be made by the teaching assistant.
- It is strongly recommended that groups select articles that include both source code and experimental data to facilitate thorough analysis and understanding.

The structure of the report should adhere to the following format:

- *Group Information*: This section should provide detailed information about the members comprising the team, including their respective roles and contributions to the project.

- *Current Status for Each Requirement*: An overview of the current progress status pertaining to each assigned requirement should be provided, including any challenges encountered and strategies employed to address them.
- *Introduction*
  - Overview of the Chosen Topic: This subsection should provide a comprehensive overview of the selected topic, elucidating its significance and relevance within the field.
  - Motivation for Research: A detailed explanation of the motivations driving the research on the chosen topic should be provided, outlining the potential impact and implications of the study.
  - Scientific Significance and Real-World Applications: This subsection should delve into the scientific significance of the chosen topic, as well as its practical applications in real-world scenarios.
- *Preliminaries and Backgrounds*
  - Notation and Definitions: Define any relevant notation and terminology essential for understanding the subsequent sections of the report.
  - Problem Statements: Clearly articulate the problem statements addressed by the chosen topic, providing context and background information.
  - General Frameworks: Present general frameworks or theoretical foundations relevant to the chosen topic.
  - Challenges and Limitations: Identify and discuss the challenges and limitations inherent in the chosen topic, highlighting areas for further research and improvement.
- *Related Works*: Review existing literature pertinent to the chosen topic, providing an overview of relevant studies, methodologies, and findings.
- *Methodology*
  - Input and Output: Define the input and output of the chosen methodology, elucidating the process through which data is processed and analyzed.
  - Advantages and Disadvantages: Evaluate the advantages and disadvantages of the selected methodology, considering its efficacy, scalability, and applicability to real-world scenarios.
- *Experimental & Evaluations*: Present the experimental procedures undertaken to evaluate the chosen methodology, along with the corresponding results and findings.
- *Conclusion*: Provide a comprehensive conclusion summarizing the key findings and insights gleaned from the study, as well as potential avenues for future research and development.

The structure of the presentation slide should encompass the following components:

- *Introduction*: Provide an introductory overview of the presentation, outlining the key objectives and topics to be covered.
- *Overview*: Contextualize the research by discussing its relevance, motivation, and significance within the field.
- *Problem statements*: Clearly articulate the problem statements addressed by the research, providing background information and context.
- *Related works*: Review existing literature relevant to the research topic, highlighting key studies, methodologies, and findings.
- *Methodology*: Present the methodology employed in the research, outlining the approach, techniques, and procedures utilized.

Table 2: Evaluation Criteria for Task 01

Criteria	Mark
Full presentation with three assignment topics/ techniques.	45%
Clear and detailed examples for each assignment topics/ techniques.	30%
Report: The report presents a logical and clear layout and format.	15%
Implementation.	10%
Bonus points for working and present more with two topics/ techniques.	20%
Total	**120%**

Table 3: Evaluation Criteria for Task 02

Criteria	Mark
Seminar (Midterm + Final)	50%
Report: The report presents a logical and clear layout and format. Analysis and study in deeply.	40%
Overall comprehension of the submitted source code.	10%
Bonus points for working and present the second paper.	20%
Total	**120%**

- *Experimental & Evaluations*: Present the experimental procedures and results obtained during the research, along with an analysis of their significance and implications.
- *Conclusion*: Provide a conclusive summary of the presentation, highlighting key findings, insights, and implications for future research.

Each group is expected to deliver a presentation lasting between 30 to 45 minutes, with all members actively participating. Presentation slides should incorporate illustrative elements such as figures, diagrams, and tables to enhance clarity and comprehension.

7   Evaluation Criteria

Your assignment will be evaluated based on the evaluation criteria which are presented in Table 2 and Table 3.

8   What to Submit

You must submit:

- ❏ **Docs Folder**: This folder should contain your report files in .doc, .docx, or .pdf formats, with a strong recommendation for .pdf to ensure compatibility and preservation of formatting. Your report should cover several key areas:
  - **Your Information**: Include your group name and student IDs.
  - **Requirement Fulfillment**: Discuss how fully each project requirement has been met.

- **Study algorithm**: Provide thorough explanations of the algorithms used, include running examples, and offer commentary on the code.
- **Presentation Style**: Aim for clarity in your report, using illustrations where helpful to convey your points effectively.

❏ **Source_Codes Folder**: This directory should house all the source code for your project. It will primarily contain Jupyter notebooks and Python scripts developed by your team. If you have code in languages other than Python, please include clear instructions for its use.

By organizing your submission in this manner, you'll help ensure that your work is clearly presented, easily navigable, and thoroughly documented, reflecting the depth and breadth of your project efforts.