

DATS369 Machine Learning with Graphs Homework 3

Due: April 21, 2025

100 Points

Instructions

- **Collaboration policy:** Homework must be done individually, except where otherwise noted in the assignments. “Individually” means each student must hand in their own answers, and you must write and use your own code in the programming parts of the assignment. It is acceptable for you to collaborate in figuring out answers and to help each other solve the problems, and you must list the names of students you discussed this with. We will assume that, as participants in an undergraduate course, you will be taking the responsibility to make sure you personally understand the solution to any work arising from such collaboration.
- **Online submission:** You must submit your solutions online on the course Brightspace site. You need to submit (1) a PDF that contains the solutions to all questions (2) `x.py` or `x.ipynb` files for the programming questions. We recommend that you type the solution (e.g., using L^AT_EX or Word), but we will accept scanned/pictured solutions as well (clarity matters).
- **Generative AI Policy:** You are free to use any generative AI, but you are required to document the usage: which AI do you use, and what’s the query to the AI. You are responsible for checking the correctness.
- **Computational Resource:** You are encouraged to use “Google Collab” (<https://colab.research.google.com/>) and Greene HPC during this class, which is free and easy to install running environment.
- **Late Policy:** Due to the spring festival, we give extra time for this homework. **No late submission is allowed.**

Clarification

This homework serves as your **final project proposal**. The only requirement is that your project must be **graph-related**. You are free to choose from a wide range of topics, including but not limited to: **graph-related predictive tasks**: classification and link prediction; **graph-related generative tasks**: text generation and image generation; **Graph-enhanced LLM/VLMs**: retrieval-augmented generation and chain-of-thought reasoning.

To receive full credits, your report should include the following key sections:

1. *(20 points)* **Introduction.** Clearly identify a novel **graph-related research topic** that genuinely interests you. In this section, describe: 1) The problem you are addressing; 2) Why this problem is important; 3) How current state-of-the-art methods attempt to solve it. **Please discuss your topic with me to ensure it is appropriate for the final project!**
2. *(30 points)* **Baseline Explanation.** Based on your chosen topic and related literature, select a representative baseline algorithm and explain its technical details. This should include: 1) Problem formulation and notations; 2) Model architecture and design choices; 3) Training objectives and loss functions. **Please consult with me to confirm that your baseline is sufficiently challenging. Ideally, it should combine Graph-based methods with LLMs or VLMs.**

3. *(40 points)* **Baseline Reproduction.** Implement and reproduce the main results of your selected baseline method. In your report, be sure to: 1) Describe the experimental setup; 2) Present your results clearly; 3) Comment on how well your reproduction matches the original work.
4. *(10 points)* **Future Plan.** Outline your plans for improving upon the baseline in the final project phase. Your general objective should be to improve the efficacy (e.g., performance or efficiency) of the baseline model.