CSCI-4280/6280: Graph Mining

`https://www.cs.rpi.edu/~slotag/classes/SP26m/index.html`

Prof. Slota, `slotag@rpi.edu`
Office Hours: Monday/Thursday 2-3pm in 317 Lally and by appointment
`https://rensselaer.webex.com/meet/slotag`

Class Hours: 12-1:50pm Monday/Thursday in Jonsson 3210

1 Course Description

This class is an introduction to graph processing and mining. Students will learn about research and analytical challenges related to the study of real-world graphs. Course topics may include, but are not limited to, the following:

- General computational graph processing
- Connectivity and centrality algorithms
- Recommender systems
- Link prediction
- Community detection
- Graph neural networks
- Subgraph mining, motif finding, anomaly detection
- Random walks
- Random graph models
- Graph ordering, compression, partitioning
- Linear algebra on graphs

1.1 Prerequisites

The only formal prerequisite for the course is CSCI-2300, Introduction to Algorithms. However, students should also be comfortable in programming with Python and C/C++. We're going to be using Python+NetworkX for a substantial portion of the class.

1.2 Course Resources

There is no official textbook for this course. Instead, we will be utilizing a wide variety of textbooks and papers found online. We will also extensively utilize a number of available repositories housing real-world graphical data. Check the website for an updated list of resources.

2 Course Schedule

Class will meet every Monday and Thursday at 12:00pm in Jonsson 3210, with the following exceptions:

Jan 19: **No** class – MLK Day
Feb 17: **Yes** class – Monday schedule

There will be no exams in this course. However, we will have scheduled group project report updates at regular intervals throughout the semester. The currently-scheduled dates are below (**may change**):

Mid-February: Proposal presentation
Mid-March: Update presentation
End of April: Final presentation
April 29: Final report due

Note that, due to the size of the class, it may be necessary to have presentations spread across two class days. To encourage you to show up for both days and participate in any discussion, groups will be randomly selected on each day. E.g., you'll need to show up prepared to present on both days (though I can't stop you from skipping day 2 if you presented on day 1).

For an up-to-date schedule with class notes and content, check the website.

3 Coursework and Grading Policies

Homework policy: Homework will comprise 40% of the course grade. There will be approximately 2-4 homeworks throughout the semester, on an approximate monthly basis. You will be expected to present some subset of your homework solutions at various points in the semester. Some homeworks may also be assigned as groupwork, due to the challenging nature of certain problems.

We will occasionally work on certain homework problems in class, so be sure to bring your laptop. Collaboration is allowed on individual homeworks, **but you must still complete all problems on your own**. Homework submissions will be collected in Submitty. You will have 5 total late days to use through the semester with a max of 2 on a given assignment. No

other late homeworks will be accepted without PRIOR approval (i.e., before the deadline) of the instructor.

Paper Presentation (4280): You will be required to read and present on two separate papers on given dates during this semester. You'll be able to select your desired papers and dates from a list given on the website, with a first-come first-served policy. These presentations will be worth 10% of your final grade (5% for each). You will be expected to prepare and give a 20 minute presentation for each assigned paper. Specific expectations will be discussed in class.

Literary Review (6280): Students taking the graduate class will instead need to perform a more comprehensive literary review of at least 5 papers relevant to their graduate research project or a separate project specific to this class. This will be prepared as a 20-30 minute presentation to the class as well as a written document to be submitted. Specific expectations will be discussed in class.

Project policy: There will be one course project to be worked on throughout the semester. Projects can be done in teams of up to 2 students each. We will discuss possible topics in class. The project will comprise the other 50% of the course grade. It will be further divided down into grades for a project proposal presentation (10%), a status report update presentation (10%), a final presentation (10%), and a final project report and submission (20%). Expectations for each of these will be discussed in class. Generally, the update presentations will be kept relatively short - under 15 minutes.

AI/LLM Policy: In the sake of fairness and equity, you are free to use AI tools to assist with your assignments. However, as mentioned above, you should still work through and understand the problem solutions on your own. Simply copy+pasting a solution is considered an academic integrity violation.

Grade Modifiers Policy: Grade modifiers will be used in this class. You can expect to earn a B- if your score is greater than 79.5 and less than 83, B if your score is greater than 83 and less than 86, B+ if your score is greater than 86 and less than 89.5. The similar modifier points occur for the A, C and D ranges except that there is no A+ nor D- under the RPI Grade Modifier Policy. **Requests for grade changes will be ignored, unless there was an identifiable error on my part**.

Curve Policy: Curves will be applied to final grades up to the discretion of the instructor. However, it is unlikely that we'll need to use a curve for the class.

4 Academic Integrity

Collaboration is allowed for homeworks, but you are **required to complete every problem on your own**. You can't just share and copy answers, but you can share ideas and approaches. This means that turning in an identical copy of your classwork's work or otherwise misrepresenting another person's work as your own will be considered an academic

integrity violation. Don't be lazy.

The evaluation of student performance is a service provided by Rensselaer. Attempts to undermine this service lower Rensselaer's reputation. Therefore, it is essential that academic honesty be preserved. Students who violate the spirit or letter of these rules are subject to penalties according to the principles outlined in the Rensselaer Handbook of Student Rights and Responsibilities:

`https://rpi.app.box.com/file/1632872852550?s=p1b39vybm914b5cao4csrgl84bfrrh2l`

In this class, first-time violations to the above policies will result in a zero on the homework/quiz/exam with which the violation occurred. Any subsequent violations will result in failure of the course. These punishments are up to the discretion of the instructor, but, generally speaking, I probably won't be particularly forgiving for egregious violations.

5 Policies for a Healthy Learning Environment

We are all different. The instructor will do his best to create an open and welcoming learning environment for all students. In this class, there is a zero-tolerance policy for harassment, bullying, racism, sexism, or any other behaviors that negatively harm this open and welcoming environment. The instructor reserves the right to immediately fail you, remove the course from your registration, banish you from the class forever, and refer you to the school for further disciplinary action should you violate this policy.

6 Disability Service for Students

Please contact me as soon as possible if you require any accommodations for the course.

7 Learning Outcomes

At the end of this course, you will:

- be able to implement and run graph analytical algorithms on real-world data
- understand basic approaches for **social, web, and biological network analysis**
- learn about ongoing research challenges with regards to mining rich, complex, graphical datasets