CSCI-4964/6964: Graph Mining
www.cs.rpi.edu/~slotag/classes/SP20t/index.html

Prof. George M. Slota
gmslota@gmail.com
Office Hours: Monday/Wednesday at 4-6pm in 317 Lally

Class Hours: 12-1:50pm M/Th in 3051 Low

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
- 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

Classes will meet every Monday and Thursday at 12:00pm in 3051 Low, with the following exceptions:

Jan 20: **No** class - MLK Day
Feb 18: **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**):

Mar 5: Proposal presentation
April 2: Update presentation
April 27: Final presentation
April 29: Final report due

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

3 Coursework and Grading Policies

Homework policy: Homework will comprise 50% of the course grade. There will be approximately 5-6 homeworks throughout the semester, on an approximate bi-weekly basis. We will work on certain homework problems in class, so be sure to bring your laptop. Collaboration is allowed on homeworks, **but you must still complete all problems on your own**. Homework submissions will be collected in Submitty. No extensions or late homework will be accepted outside of verifiable and approved circumstances, as we will go over homework solutions in-class.

Project policy: There will be one course project to be worked on primarily in the second half of the semester. Projects can be done in teams of up to 4 students each. We'll 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 (5%), a status report update presentation (10%), a final presentation (15%), and a final project report and submission (20%). Expectations for each of these will be discussed in class.

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 with homeworks and projects. However, for homeworks, you are still **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:

`https://info.rpi.edu/sites/default/files/Handbook-of-Student-Rights-and-Responsibilities-Rev-August-29-`

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.

5 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