EECS6414 Data Analytics and Visualization

Winter 2024

Course Website
www.eecs.yorku.ca/~papaggel/courses/eecs6414/

Course Description
Data analytics and visualization is an emerging discipline of immense importance to any data-driven organization. This is a project-focused course that provides students with knowledge on tools for data mining and visualization and practical experience working with data mining and machine learning algorithms for analysis of very large amounts of data. It also focuses on methods and models for efficient communication of data results through data visualization.

Topics
Topics include:

- basic graph theory
- network measurements
- network models
- community detection
- link analysis & prediction
- cascading behavior in networks
- value of visualization
- visualization of multidimensional data
- visualization of networks
- tools for network analysis and data visualization

Instructor
Manos Papagelis
Email: papaggel@eecs.yorku.ca, papaggel@gmail.com
Website: http://www.eecs.yorku.ca/ ~papaggel

Teaching Assistant
Jing Li (jliellen@yorku.ca)

Class Hours
Lectures: Wed 13:00pm-16:00pm at BRG 213 (Bergeron Building)
Office Hours: Thu 16:00-17:00 at LAS3050 (or right after class)

Class Attendance
Attendance of lectures is expected but not required.

Prerequisite Courses
There are no course prerequisites for this course; students are expected to be competent in:

- Algorithms (basic data structures, dynamic programming, …)
- Basic probability & linear algebra (moments, typical distributions, MLE, …)
- Programming (your choice, but Python/C++/Java will be very useful)

If you don't satisfy these, you need to talk with the instructor in the first week of classes to see whether you may remain in the course.

Textbooks
The course will rely mainly on the following suggested textbooks:

- Mining of Massive Datasets, 3rd Edition (2020) by Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman (freely available online)l
- Networks, Crowds, and Markets: Reasoning About a Highly Connected World (2010) by David Easley, Jon Kleinberg (freely available online).

- Book title: Introduction to Data Mining, Second Edition (2017) by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar
- Social Media Mining: An Introduction (2014) by Reza Zafarani, Mohammad Ali Abbasi, Huan Liu (freely available online).
- The Visual Display of Quantitative Information, 2nd Edition (2001) by Edward R. Tufte
- Envisioning Information (1990) by Edward R. Tufte
- Interactive Data Visualization for the Web, 2nd Edition (2017) by Scott Murray

In addition, a number of recent research papers in the area of data analytics and visualization will be distributed in every iteration of the class.

Communication

The following are the communication tools for the class.

- ***eClass/Moodle***: https://eclass.yorku.ca/course/view.php?id=90853
- ***Course Website***: https://www.eecs.yorku.ca/~papaggel/courses/eecs6414/
  All class material will be available on the course web site; check regularly. The page also has a link to a discussion board. We are using Piazza.
- ***QA Forum (Piazza)***: https://piazza.com/yorku.ca/winter2024/eecs6414
  Instead of a discussion board, we are using Piazza, a free Q&A platform. Piazza can get you fast, accurate response to your questions – but it only works if everyone participates! We will also use Piazza to post announcements and updates, so both the website and Piazza is required reading.
  Note: You will need to sign up with your school email, ending in *yorku.ca*. If you do not have a school email address, please contact your instructor and request to be enrolled with your personal email.
- ***Email***: Please **use email only for personal issues** and the QA forum to ask course-related questions. Include "eecs6414" in all email subject lines to ensure your message is correctly filtered and filed. An informative subject line like "eecs6414: Question related to X" really helps. I try to respond to email frequently. However, due to volume, it may take longer, especially on weekends and near due dates.

Grading Policy

The evaluation of the course will be based on projects done by teams (of up to three)

Milestone	Weight
Project proposal	10%
Project midterm report	20%
Project midterm in-class presentation	10%
Project final report	40%
Project final in-class presentation	20%

Final Examination

There is no final examination.

Working with a Partner

You have the option of partnering with other (currently enrolled) students for your project, and we encourage you to do so. The ability to work effectively in a team will be very important in your career, and that involves many skills beyond the purely technical aspect of creating working code. You may choose your own partner(s). If you do have a partner, submit only a single copy of your work. Jointly submitted assignments will be graded in the usual way and all partners will receive the same mark. Working with a partner has the potential to lighten your workload or to increase it, depending on how well you work together. Be aware that simply dividing the work and assembling your separate

pieces at the end is a poor strategy for completing successful assignments. And of course, you are responsible for learning the course material underlying all parts of the assignments. You will have the most success if you truly work together.

Assignment Policies
Here assignment refers to project milestones. You must make sure that all your assignments are running and are sufficiently documented. Code that doesn't compile, fails to run or lacks documentation, will be marked as not working.

Late Work Policy
The late policy is strict. All assignments and project deliverable will be submitted electronically. Late submissions will be handled based on a system of **"one grace day"** per deliverable, as follows: you can use one grace day for each assignment submission or project deliverable without requesting permission; one grace day is 24 hours. If an assignment is due at 11:59 p.m. on a Friday, then an assignment handed in by 11:59 p.m. on Saturday uses one grace day. The intention of a grace day is for use in emergencies (e.g., system failure or illness). Assignments submitted after the due date plus one grace day will receive a grade of 0. If you are at risk of missing a deadline due to a busy week, rather than requesting extra grace days you should hand in a working (and tested) version of a simpler program. In the event of an illness or other catastrophe, get proper documentation (e.g., medical certificate), and contact me (by email or in person) as soon as possible. Do not wait until the due date has passed. It is easier to make alternate arrangements before the due date.

Remarking
Here assignment refers to project milestones. If you feel an error was made in marking an assignment or test please submit a remark request. Requests for remarking must be submitted using a university remarking request form explaining what your concern is **no later than a week after** the assignment (or test) has been returned back.

Policy on Generative AI tools and Technologies (ChatGPT, etc.)
The use of generative AI tools and technologies, such as ChatGPT, to create content is permitted but must be fully disclosed in the report. For example, the authors could include the following statement in the Acknowledgements section of the Work: "ChatGPT was utilized to generate sections of this report, including text, tables, graphs, code, data, citations, etc.)." If you are uncertain about the need to disclose the use of a particular tool, err on the side of caution, and include a disclosure. Basic word processing systems that recommend and insert replacement text, perform spelling or grammar checks and corrections, or systems that do language translations are to be considered exceptions to this disclosure requirement and are generally permitted and need not be disclosed in the report. As the line between Generative AI tools and basic word processing systems like MS-Word or Grammarly becomes blurred, this policy will be updated.

Academic Offenses
All of the work you submit must be done by you and your work must not be submitted by someone else. Plagiarism is academic fraud and is taken very seriously. The department uses software that compares programs for evidence of similar code. Please read the Rules and Regulations from the York University's Academic Integrity and the York University's Senate Policy on Academic Honesty documents.

Accessibility Needs
If you require accommodations for a disability, or have any accessibility concerns about the course, the classroom or course materials, please contact York University's Counselling & Disability Services.