EECS4414 Information Networks (Cross-listed as EECS5414)

Fall 2021

Course Website
https://www.eecs.yorku.ca/~papaggel/courses/eecs4414/

Course Description
Information networks are effective representations of pairwise relationships between objects. Examples include technological networks (e.g., World Wide Web), online social networks (e.g., Facebook), biological networks (e.g., Protein-to-Protein interactions), and more. The study of information networks is an emerging discipline of immense importance that combines graph theory, probability and statistics, data mining and analysis, and computational social science. This course provides students with both theoretical knowledge and practical experience of the field by covering models and algorithms of information networks and their basic properties. In addition, analysis of information networks provides the means to explore large, complex data coming from vastly diverse sources and to inform computational problems and better decisions.

Topics
Topics include:

- basic graph theory
- network measurements
- network models
- link analysis & link prediction
- network ties
- community detection
- graph partitioning
- information cascades & epidemics
- graph mining
- machine learning for graphs
- connections to problems in the social sciences and economics

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

Teaching Assistant
Tilemachos Pechlivanoglou (tipech@eecs.yorku.ca)

Live-class Hours
Lectures: *Mon*, 18:00 – 21:00 online (on Zoom); some weeks we will use the 3rd hour for a tutorial.
Office Hours: *Mon*, 13:00 – 14:00 online (on Zoom).

Class Attendance
Attendance of lectures is expected but not required.

Prerequisite Courses
The course prerequisites for this course are:

- EECS-3421: Introduction to Database Systems
- EECS-3101: Design and Analysis of Algorithms
- MATH-2030: Elementary Probability
- General prerequisites

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
There is no single text for this course. The course will rely mainly on the following suggested textbooks:

- Networks, Crowds, and Markets by David Easley, Jon Kleinberg (freely available online), 2010
- Networks: An introduction by Mark Newman, 2010
- Social and Economic Networks by Matthew O. Jackson, 2010
- Mining of Massive Datasets, 2nd Edition by Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman (freely available online), 2014
- Social Media Mining by Reza Zafarani, Mohammad Ali Abbasi, Huan Liu (freely available online), 2014

The course will also rely on the following reference notes/articles/reviews:

- The Structure and Function of Complex Networks (by M. E. Newman), 2003
- Structure and Dynamics of Information in Networks (by David Kempe), 2011

In addition, a number of research papers on graphs and information networks 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/eclass/course/view.php?id=48747
- ***Course Website***: https://www.eecs.yorku.ca/~papaggel/courses/eecs4414/
  All class material (except for video recordings) will be available on the course web site; check regularly. The page has also a link to a discussion board. We are using Piazza.
- ***QA Forum (Piazza)***: https://piazza.com/yorku.ca/fall2021/eecs44145414
  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 "eecs4414" in all email subject lines to ensure your message is correctly filtered and filed. An informative subject line like "eecs4414: 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

Work	Weight	Comment
2 Assignments	30%	A1: 10% and A2: 20%
Project	40%	Large project consisting of research proposal, milestone report, final presentation, and final report and code
Final Exam	30%	You must get >=40% to pass the course

Final Examination
A written final exam will be given between Dec 9-23 (to be determined during the term).

Working with a Partner for Assignments (up to 2 students) and the Project (up to 4 students)
You have the option of partnering with one other (currently enrolled) student for your assignments (**teams of up to 2 students**) and three others for your project (**teams of up to 4 students**), 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, and it need not be the same person for each assignment. If you do have a partner for an assignment, submit **only a single copy of your work**. Jointly submitted assignments will be graded in the usual way and both 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 success. 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 and Project Policies

You must make sure that all your assignments and project deliverable 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 a 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 or test day.

Assignments are submitted electronically; you must follow the submission instructions and filename conventions exactly. If you do not, you will most likely lose substantial marks on the assignment. If you find you have submitted the wrong file or omitted a file, please notify your instructor as soon as possible.

Remarking

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.

Academic Offenses

All 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

York University is committed to accessibility. 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.

Tentative Schedule

A tentative schedule of topics to be covered appears below. This is subject and likely to change.

Topic
Lecture 1: Introduction Introduction, administrivia, introduction to main problems about networks, basic mathematical concepts, bow-tie structure of the Web.
Lecture 2: Network Measurements Degree distributions, shortest paths, clustering coefficient, measuring power-laws.
Lecture 3: Network Models Erdos-Renyi random graph model, small-world model, configuration model, power-law distributions, scale-free networks, the anatomy of the long-tail, preferential attachment model, macroscopic evolution of networks, forest-fire model.
Lecture 4: Link Analysis and Node Importance in Networks: HITS and PageRank Web search, Hubs and Authorities (HITS), PageRank, topic-sensitive PageRank, personalised PageRank.
Lecture 5: Link Prediction in Networks Link prediction, neighborhood-based prediction methods, node proximity based prediction methods, supervised learning models, Facebook's "PYMK" algorithm, Twitter's "WtF" algorithm.
Lecture 6: Strength of Weak Ties & Community Structure in Networks Strength of weak ties, structural holes, network communities, community detection, Girvan-Newman algorithm, modularity, modularity optimization.
Lecture 7: Network Community Detection: Graph Cuts and Spectral Clustering Graph partitioning, graph cuts, conductance, spectral graph theory, spectral graph clustering.
Lecture 8: Overlapping Communities Overlapping communities, cliques, clique percolation method, modeling networks with communities, community-affiliation graph model.
Lecture 9: Cascading Behavior: Decision Based Models of Cascades Spreading through networks, Granovetter's model of collective action, decision based model of diffusion, game theoretic model of cascades.
Lecture 10: Cascading Behavior: Probabilistic Models of Information Flow Epidemic model based on trees, models of disease spreading (SIR, SIS, SIRS), independent cascade model, modeling interactions between contagions.
Lecture 11: Advanced Topics: Machine Learning with Graphs Network representation learning, graph convolutional networks (GNNs), knowledge graphs (KG)
Lecture 12: Team Project Presentations In-class team project presentations

Tentative Assignment Schedule (30%)

Assignment	Weight	Posting Date	Due Date	Topic
1st	10%	Fri, Oct 1	Fri, Oct 15	Network Models & Measurements
2nd	20%	Fri, Oct 22	Fri, Nov 12	Link Analysis & Prediction; Community Detection

Tentative Project Schedule (40%)

Milestone	Project Milestone Weight	Due Date
Project proposal	20%	Fri, Oct 8
Project progress report	20%	Fri, Nov 5
Project final report & code	50%	Fri, Dec 3
Project In-class Presentation	10%	Mon, Dec 6