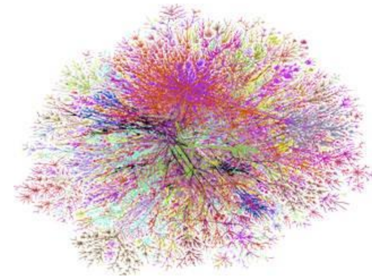
ECE 5260 – ORIE 5735

GRAPH-BASED DATA SCIENCE FOR NETWORKED SYSTEMS



Complex networks are often the source of high-dimensional data. The goal of this course is to introduce structural and computational models for data that are indexed by the irregular support of a graph. The graph represents the network that couples the dynamics of many agents, or it can be a more abstract Bayesian graphical model that explains how observations are conditionally dependent. The interest in these models spans many fields.

OBJECTIVES:

1. (Module 1) - **Networks and graph models** - introduce models that capture different **structural characteristics of networks** and network-scientific **metrics to compare them**
2. (Module 2) - **Network dynamics and data models** - introduce mathematical **models to explain and simulate how data originate from network dynamics** (e.g. dynamical models for social, biological, financial networks, and information networks, and cyber-physical infrastructures like packet-switched networks, transportation networks, energy, water etc.)
3. (Module 3 and 4) - model inference problems on these types of data process the observations through algorithms that **leverage the network description namely**
    1. **Graph Signal Processing** - the emerging field of **graph signal processing**, a theory that generalizes digital and image processing to graph signals.
    2. **Graphical models** - Bayesian networks, modeling conditional independence in multivariate data models, which helps deriving optimum Bayesian inference algorithms for detection and estimation.

OUTCOMES

The outcomes of this course are as follows. Students will learn:

1. to analyze graph data as networks and understand their structural features
2. models of network dynamics that occur in science and engineering that utilize the algebra of networks
3. techniques to analyze data that result from network dynamics in Graph Signal Processing
4. Bayesian techniques that apply to random events whose dependencies can be expressed through a graphical model.

INSTRUCTION

- **Instructor:** Prof. Anna Scaglione, office: Bloomberg 265
- **Teaching Assistant:** Dr. Tong Wu
- **Course schedule:** M-W, 2:45pm - 4:00pm, Bloomberg 91

- Office hours: T, 3:00pm - 5:00pm on zoom

Requirements

The main background requirements are linear algebra and knowledge of probability theory. You can check http://snap.stanford.edu/class/cs224w-2017/recitation/proof_techniques.pdf . Knowledge pf basic time-series/digital signals analysis is useful but not necessary.

Reading material

- William Hamilton "Graph Representation Learning", Morgan and Claypool.
- Eric D. Kolaczyk "Statistical Analysis of Network Data: Methods and Models (Springer Series in Statistics) 2009th Edition"
- Albert-László Barabási "Network Science"
- For Networkx in addition to the tutorial I will also draw from: ***A First Course in Network Science:*** https://cambridgeuniversitypress.github.io/FirstCourseNetworkScience/

Course modules

- (Module 1) - For the first part of the course that focus on modeling complex networks as graphs and introducing metrics and algebra to unveil their properties.
- (Module 2) - For the second part I will supplement the books above with notes to detail the dynamics of various sample networks and how to simulate the data.
- (Module 3) - For the part on Graph Signal Processing I will use Hamilton's book but also refer to tutorials, such as:
  - Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., & Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. Proceedings of the IEEE, 106(5), 808-828. https://ieeexplore.ieee.org/abstract/document/8347162/
  - Abadal, Sergi, Akshay Jain, Robert Guirado, Jorge López-Alonso, and Eduard Alarcón. "Computing graph neural networks: A survey from algorithms to accelerators." *ACM Computing Surveys (CSUR)* 54, no. 9 (2021): 1-38. https://dl.acm.org/doi/abs/10.1145/3477141
  - https://github.chttps://colab.research.google.com/github/deepmind/graph_nets/blob/master/graph_nets/demos/graph_nets_basics.ipynb
- (Module 4) For the part on Bayesian Networks I will use the main book and additions from Jordan, Michael I. "Graphical models." *Statistical science* 19.1 (2004): 140-155.
  - Python code: https://pgmpy.org/

Specific topics

- **Module 1 - Networks and graph models**
  - Introduction to the course
  - Graph theory primer
  - Handling networks in python
  - Analysis of graphs
    - comparative graph metrics and feature extraction
    - communities and their detection
  - Random graphs models
- **Module 2 - Network dynamics and data models**

    - Deterministic and stochastic dynamics
    - Markov models
    - Examples
        - Consensus models
        - Epidemic models
        - Social dynamics and opinion diffusion
        - Financial networks dynamics
        - Examples from biological networks
    - Infrastructures
        - information networks and the web
        - Physical systems: electric grid, gas grid, transportation
    - Examples of graphical models
- **Module 3 - Graph Signal Processing**
    - Review of DSP basics
    - The graph shift operator and Graph Filters
    - Interpretation of data from network dynamics as graph signals
    - The Graph Fourier Transform
    - Graph Neural Networks
- **Module 4 - Graphical Models**
    - Bayesian networks and factor graphs
    - Max product/ max sum algorithms

Subject to Change Notice
All material, assignments, and deadlines are subject to change with prior notice. It is your responsibility to stay in touch with your instructor, review the course site regularly, or communicate with other students, to adjust as needed if assignments or due dates change.

Assignments

The assignments will vary for PhD versus MS students. For MS students they will consist primarily in applying the models provided in class for simulation, and numerical implementation of algorithms to analyze network data using Python. PhD students will also have some graph theoretic problems to solve and will be asked also to advance modeling of graph-based data. The assignment will be available online and should be submitted electronically. There will be typically weekly assignment in the course, and at least 2 of them will be biweekly, when more time consuming (you can consider them like midterms).

- Weekly homework - 30% of the grade
- Midterms - 35% of the grade
- Final Project - 35% of the grade

Due dates: *Unless otherwise indicated on Canvas, the weekly assignments are due a week from when they are assigned and the biweekly ones two weeks after they are assigned.*

Working the Final Project

The final project is in teams and students can form their teams or will be assigned a partner. Partners will individually submit the assignment they worked on together indicating 1) who their partner is and 2) the part that they took responsibility for (cannot be the same for both partners). The partners will receive a grade that is the average of a grade for the group performance and an individual grade, based the grade for the part they lead.

Grades scale

The grade scale used to grade is from 0 to 4, indicating

- 0 = nothing to work with for assigning a grade
- 1= poor
- 2= fair
- 3= good
- 4 = excellent

The final grade average will be rescaled in the range [0,100] and the following thresholds will be used for assigning the letter grade:

Letter Grade	Percentage	Points Range	Letter Grade	Percentage
A+	100 – 96.0%	M– 0.96*M	C+	72.9% -67%
A	95.9 – 93.0%	0.969*M–0.92*M	C	67% -56%
A-	92.9 – 89%	Etc.	D	55% – 50%
B+	88.9 – 85.0%		E	49%–43%
B	84.9 – 80.0%		F	
B-	79.9 – 73%			

Late or Missed Assignments

Notify the instructor BEFORE an assignment is due if an urgent situation arises and the assignment will not be submitted on time. You can miss one of the homework without impact on your grade because of an emergency. You can gain 10% over the grade for early submissions (24h before the deadline) and loose 10% for every day the assignment is late for up to 2 days.

Student Success

To be successful:

- read announcements
- read and respond to course email messages as needed
- complete assignments by the due dates specified
- communicate regularly with your instructor and peers
- create a study and/or assignment schedule to stay on track

Attendance/Participation

Preparation for class means reading the assigned readings, practice problems and reviewing all information required for that week. Attendance for students enrolled in this course means in on a regular basis attending classes, logging on to Canvas to check for assignments and participating in all the pertinent activities that are posted in the course.

Attendance feedback to the instructor is also part of the course attendance

Course Time Commitment

The course requires you to spend time preparing and completing assignments. A three-credit course requires 135 hours of student work. Therefore, expect to spend approximately 7-8 hours a week preparing for and actively participating in this course.

Students with Disabilities

Please make sure you notify the instructor, and Students for Disabilities serves at the beginning of the course regarding special needs: http://sds.cornell.edu.

Communicating With the Instructor

This course uses a "three before me" policy regarding student to faculty communications. When questions arise, please remember to check these three sources for an answer before asking me to reply to your individual questions:

1. Course syllabus
2. Announcements in Canvas
3. Office hours