CSS 692 Spring 2022 Syllabus: Social Network Analysis

Instructor: Eduardo López
Department: Computational and Data Sciences
email: elopez22@gmu.edu
Zoom class area: Click here
Class time: Thursdays 16:30

Overview: This class focuses on the foundations, applications, and algorithms of social network analysis. For a relatively long period of time, social networks have been used as a quantitative tool to understand the social. Considerable progress was made in this area from the time of Moreno's sociograms ($\sim$ 1930) until a more recent wave of interest started around 1998 with the work of Watts and Strogatz focusing on the onset of the small-world property. This latter work coincided with the explosion of data that came with the information age, and propelled social network analysis, along with other technical areas that used or adopted networks, to a prime status in the realm of data analysis. The course will basically be organized into blocks. First, we will formalize the concept of a network. For this, we will talk about the basic ingredients that form a network (their nodes and links), some of the basic quantities that one cares about when studying networks, and a set of basic computational tools that can be used to do this basic analysis. The next block of the course will concentrate on introducing features into the elements of the network. These features, like node characteristics or weights along the links, mimic the real phenomenology that accompanies networks. When such features are introduced, one can begin to explore new questions relevant to social systems. In the final block, we aim to discuss network models and how these address a new set of questions that revolve around the features that may (or may not) lead to special social networks. The class will rely on a combination of tools, predominantly some formal concepts of networks presented in terms of algebra and linear algebra, computational tools for the creation and analysis of real-world networks, and some statistical tools to extract meaning from network models.

Flip Class Format: Students should be aware that the intended format of this class is as a *flip classroom*. Material will be provided well in advance of the class, and the class period will be used for discussion of the readings. As the instructor, I will begin classes by discussing an overview of the class reading and then nudging the discussion along on the basis of the readings. **After some sessions, I will be happy to discuss with the class if the format is generally working, and whether or not changes are needed.**

Office hours: These will be offered by appointment, over zoom or telephone. Please email me.

Policies and evaluation: Final score will be based on a combination of 1) Assignments (60%), 2) Final project presentation (30%), 3) Participation (10%):

Assignments: There will be three assignments with the intention of putting to practice the concepts and computational tools explained. The assignments will be discussed in class. The structure of the assignments is the following:

1. Code will be created or modified from a template to measure a given property in a network.

2. The code will be tested and applied to show correctness as well as the ability to interpret the meaning of the results.

3. The code and result will be presented in class to instructor and peers for discussion and evaluation.

Final Project: Students will pick a final project topic that involves the analysis of a real dataset that represents or can be cast into a network, or a random model that leads to non-trivial networks. At the end of the course, some sessions will be dedicated to the final presentations by the students. **Final projects will not require a written report, but those that do produce one will be eligible for an extra credit reserve of up to 10% towards their final grade in the class.** A good quality report should aim to be in the format of a research paper and/or dissertation proposal (but not exceed 4 pages). If a report is written, it should generally contain: 1) Abstract, 2) Introduction, Background, and Motivation, 3) Methods, 4) Results, 5) Discussion, 6) Conclusions, and 7) References. However if the nature of the project does not require all such sections, write as need.

About assignment and final project presentations: All presentations should be carried out efficiently and in a small period of time so that we can move through them quickly. Assignment presentations should last between 5-10 minutes per student, final project presentations around 15-20 minutes. Organization and clarity will be part of the grading.

About assignment and final project submissions: The Blackboard site for the class has an area where you will upload your code on the day it is due, before the class starts. The deadline for submitting the code is the class start time.

Participation: Coming to class and being engaged will not only help you understand and follow the class better, but it will also earn you 10% of the grade. Please take into consideration that under the distance format, your active participation in the virtual sessions and engagement in the class discussions truly enhances the experience and provides an environment that compares to in-person teaching.

Material Release: The class materials will be released periodically as opposed to all at once. I anticipate 3 release dates: Monday, January 24th (up to week 5 material), February 21st (up to week 10 material), and March 28th (remainder of material).

Assignment Release: Assignments will be released on the classes where they are indicated in the tentative schedule below, or at the agreed class session in case the schedule is modified.

Flexible Schedule: Because of current circumstances, and also taking into account the distance format of the class, the instructor and/or the class may feel the need to make adjustments to the schedule. If as a group we collective believe some changes are needed to make the class more productive, please bring it to the attention of the instructor.

Suggested Supporting Material: "Social Network Analysis" by S. Wasserman and K. Faust, Cambridge University Press (1994), "Networks: An introduction" by M.E.J. Newman, Oxford University Press (2010), "Analyzing Social Networks" by S. P. Borgatti, M. G. Everett, and J. C. Johnson, SAGE (2013).

Academic integrity: The honor code will be enforced. Mason has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow at all times are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students, give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment, ask for clarification. No grade is important enough to justify academic misconduct. Plagiarism means using the exact words, opinions, or factual information from another person without giving the person credit. Writers give credit through accepted documentation styles, such as parenthetical citation, footnotes, or endnotes. Paraphrased material must also be cited.

Disability Statement: If you have a documented learning disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with the Office of Disability Services (SUB I, Rm. 222; 993-2474; http://www.gmu.edu/student/drc/) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Tentative Schedule:

Week 1
Admin: discuss syllabus and computational tools. Basics: nodes and links, nomenclature by discipline, adjacency matrix, node degree, link lists, undirected and directed graphs, weighted graphs, network density.

Week 2
Survey of network properties: degree histogram/distribution and its types, network average degree, paths and shortest paths in networks, structural connectivity, network motifs (configurations), triangles, v-shapes, clustering. Assign first homework.

Week 3
Survey of network properties: The size of the network and the small-world property, eigenvector centrality, page-rank, betweenness (and its flavors). Milgram's experiment. Global structure of the network (clusters). First homework discussion.

Week 4
The Small-world property in depth.

Week 5
Distributions of Degree, Triangles, and Clustering. Second homework assigned.

Week 6
Distributions of Degree, Triangles, and Clustering. Second homework discussion.

Week 7
Epidemics and propagation processes in networks. Second homework discussion.

Week 8
Epidemics and propagation processes in networks. Centrality Measures

Week 9
Flavors of multi-mode networks: networks of affiliations, bipartite network. Third homework assigned.

Week 10
Survey of network properties: More on global structure, modularity, community detection, hierarchical clustering. Third homework discussion.

Week 11
Networks with attributes: Roles, structural equivalence.

Week 12
Random network models. Packages used in network analysis.

Week 13
Hypothesis testing on networks. Project presentations

Week 14
Project presentations