CDS 692 Spring 2020 (Updated COVID-19) Syllabus: Social Network Analysis

Instructor: Eduardo López
Department: Computational and Data Sciences
Office: Research Hall 254
email: elopez22@gmu.edu
Class time: Thursdays 16:30 to 19:10
Class location: Hanover Hall L002

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.

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 four 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.

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.

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.

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, nodes degree, adjacency matrix, node degree, link lists, undirected and directed graphs, weighted graphs.

Week 2
Survey of network properties: network density, network average degree, degree histogram/distribution and its types, paths and shortest paths in networks, structural connectivity. Assign first homework.

Week 3
Survey of network properties: The size of the network and the small-world property. Milgram's experiment. Global structure of the network (clusters). First homework discussion.

Week 4
Survey of network properties: network motifs (configurations), triangles, v-shapes, clustering. Assign second homework.

Week 5
Survey of network properties: Centrality measures: eigenvector centrality, page-rank, betweenness (and its flavors). Second homework discussion.

Week 6
Epidemics and propagation processes in networks.

Week 7
Survey of network properties: More on global structure, modularity, community detection, hierarchical clustering.

Week 8
Networks with attributes: link weights, multiple relations, multi-layer networks.

Week 9
Networks with attributes: Roles, structural equivalence. Assign third homework.

Week 10
Flavors of multi-mode networks: networks of affiliations, bipartite network. Third homework discussion.

Week 11
Random network models. Assign fourth homework

Week 12
Hypothesis testing on networks. Fourth homework discussion.

Week 13
More random network models. Project presentations

Week 14
Project presentations