CSS 739/CSI 709 Fall 2018 Syllabus: Exponential Random Graphs and Other Network Models for Social and Data Science

Instructor: Eduardo López
Department: Computational and Data Sciences
Office: Research Hall 254
email: elopez22@gmu.edu
Class time: Tuesdays 16:30 to 19:10
Class location: Research Hall 162

Overview: This class is about the statistical and computational foundations of network modelling, with a large portion focused on exponential random graphs (ERGMs), very common in the study of social networks. We start by explaining that all random networks are chosen from a set of possible networks called the sample space. Furthermore, depending on the problem of interest, we explain the need to assign probabilities to pick each of the networks in the sample space so that statistically tests can be performed properly. We spend time showing how to determine these probabilities computationally and, when possible, mathematically. A considerable portion of the course is concerned with learning how to create algorithms from scratch that can tackle such statistical tests. We present alternative approaches to the creation of random networks, including growth models such as preferential attachment, and configuration models. Applications to data are presented throughout the course, including examples in social networks, technological networks, and several other data-driven examples. Computation will be taught using python (not a prerequisite). This class is well suited for those with a serious interest in data analysis via networks. There is an expectation of proficiency in both computer programming, and a basic level of calculus. **For details see tentative class schedule below.**

Who is this class for: The class is an excellent introduction to social and other network modelling for graduate students in computational social science, statistics, social science, CS, data science, and other related fields. It does not assume a great deal of social science background, but presents relevant social concepts along the way. The class is a formal methodological introduction with lots of material applicable to research for those interested in the data science of social networks. It is also a useful class for those wanting to strengthen their backgrounds in statistical modelling, by applying those concepts in networks. **Applications to large data sets will be presented, with the purpose of achieving real-world job or reserach proficiency in the topic.**

Policies and evaluation: Final score will be based on a combination of 1) Assignments (45%), 2) Final project presentation (45%), 3) Participation (10%):

> **Final Project Presentations:** Each student will pick a final project topic to be discussed with instructor before the start of week 10 of term. The topic should be chosen with appropriate relevance to the course, and manageable enough to be ready to be presented by the end of the course. The project will be evaluated on the following criteria in equal proportions: 1) correct application of relevant definitions and techniques (45%, 2) clarity of presentations (45%), 3) interest and originality of the results (10%).
>
> **Assignments:** There will be assignments approximately every 2 to 3 weeks with the intention of putting to practice the concepts and computational tools explained. The assignments will be discussed in class.

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.

Material: “Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications” edited by Dean Lusher, Johan Koskinen, and Garry Robins, “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
Basics: nodes and links, nomenclature by discipline, nodes degree, adjacency matrix, link lists, undirected and directed graphs, weighted graphs.

Week 2
The graph sample space: smallest and largest networks, number of possible distinct networks, connected and disconnected networks.

Week 3
Computational basics: python language overview, tools for network computation, networkx, igraph, scipy, matplotlib, some basics of visualization.

Week 4
More network properties: degree distribution, path lengths, small-world, basics of betweenness, basics of motifs (configurations).

Week 5
Random networks basics I: picking networks with a probability, network probability spaces (ensembles), properties of network spaces.

Week 6
Random networks basics II: some computational tools, complete enumeration, computational calculation of random network properties.

Week 7
Random networks basics III: continue properties of network spaces, entropy, the maximum entropy property and its statistical relevance, maximum likelihood.

Week 8
Building a probability space (ensemble): Using entropy maximization to assign network probabilities, picking constraints, basic exponential random graphs.

Week 9
Exponential random graphs (ERGMs): broad definition, types of dependencies (network self-organization, actor attributes, exogenous factors), specific examples, basic applications, computational implications.

Week 10
Computational approaches to ERGMs: network Monte Carlo, sampling properties, basic statistical testing.

Week 11
Data analysis: data specification and research question, creation of relevant network space, applied statistical testing. Final project introduction.

Week 12
Building networks from basic principles: binomial distribution and random networks, Erdős-Rényi and Gilbert, preferential attachment, network Monte Carlo revisited.

Week 13
Optional advanced topics I: random network robustness and percolation, giant cluster, distribution of cluster sizes.

Week 14
Optional advanced topics II: bipartite networks, one-mode projections, affiliation networks, multiplex networks, triads

Week 15
Project presentations