Statistics 8810: Advanced Topics in Statistics I
Statistical Inference in Network Data
Spring 2020 (Session 2) Course Syllabus (Tentative)

Instructor: Dr. Subhadeep Paul

Lectures: 12:40 PM – 2:30 PM on Tuesdays and Thursdays in Baker Systems 136 starting Feb 26, 2020.

Office Hours: TBD at 231 Cockins Hall. You are welcome to walk in anytime during the office hours. However, if you can't make it during the specified hours, you can email me to request an appointment at a different time.

Email: paul.963@osu.edu.

Office: 231 Cockins Hall

Course Website: Important announcements, course materials and lecture notes, homework problems, computing references, and other information about the class will be posted on Carmen (carmen.osu.edu, login with your OSU name. ID).

Course Description: The course is intended to introduce the field of statistical inference in network data. The course will have a good mix of theory, methods and applications. While I assume the primary audience of the course is PhD students and senior Masters students in Statistics and Biostatistics, there will be elements which are of interest to students from other departments interested in research on network analysis.

I will provide lecture notes and reading materials (papers, review articles etc.) for each of the topics. The notes will be in the style of review. Depending upon topic, some will be in-depth while others will take a broad view.

The modules will be taught in a shuffled manner in order to simultaneously focus on application and theory/methods.

- Module 1: Introduction
  - Basic concepts and properties of graphs, motivation for network analysis,
  - Historical context of network analysis (in statistics, sociology, math, physics, computer science etc.)
  - Application areas of network data
  - Community Detection

  - Small-world and power law degrees
  - Data structures - single network, multi-layer network, time varying network, multimodal network
  - Learning tasks - community detection, link prediction, classification.

- Module 2: Erdos-Renyi random graphs and their properties. Probabilistic techniques and concentration inequalities.
- Module 3: Stochastic Block Models
  - Estimation methods - maximum likelihood, modularity, variational methods, Bayesian approaches, Spectral methods
  - Error of community detection - consistency, minimax rates, detection thresholds
  - Extensions - degree corrected, mixed membership, superimposed models, and graphons.
- Module 3: Latent Space Models
  - Modeling choices
  - Inference
  - Extensions - hyperbolic space, more complex data structures
  - Tensor methods and latent factor models
- Module 4: Dynamic and Multilayer networks
  - Multilayer networks: models and methods
  - Discrete time dynamic networks
  - Continuous time interaction networks

  Module 5: Application of network methods
  - Social network analysis
  - Network neuroscience
  - Network analysis in economics
- Module 6: Additional topics (depending upon how much time we have)
  - Exponential random graph models
  - Regression with network data
  - Hypothesis testing involving network data
  - Hypergraph learning
  - Random dot product graph

Prerequisites / Co-requisites:
As a research focused course this is open to all graduate students in statistics and biostatistics and PhD students from other disciplines. Understanding some of the topics require sufficient background in graduate level mathematical statistics, large sample theory, and probability theory. The nature of evaluation for this course (see next section) means students from diverse backgrounds can

take and do well in this course. Students who lack certain mathematical backgrounds can choose to focus on portions of the course they are more interested in and ignore some of the theoretical details. The same goes for the students who are more interested in the theory (and methods) component of the course. They can choose to not focus much on the applications.

Evaluation: There is no homework or exam for this course. The evaluation is based on a written project proposal and a final project report due at the end of the course (exact dates will be announced later). In Carmen, I will provide a list of topics to write short micro-reviews of the literature on (approximately 5-10 pages), as well as a number of research topics that I am interested in. The topics for micro-reviews will consist of statistical methods, theory as well as different application domains. The reviews are expected to be of very high quality. You are also free to suggest me a topic related to network analysis to write a review on. I will consider the topic and let you know if it is acceptable. If you take the research route, you need to declare a stated research goal (even if vague) beforehand and show progress towards achieving this goal. I don't expect you to have completed the stated objective before the semester ends, however the written report should reflect a substantial good faith effort.

For both review or research route, two reports are required. The first one is a proposal document (approx. 1 page) and the second one is a final report (5-10 pages for review, no limit for research). You are welcome to form groups (and I encourage it) of any number of members, however, I expect the effort and quality of the documents to increase linearly with the number of authors.

Computing: We will be primarily using the `R` statistical computing software for in-class demonstrations. `R` may be downloaded for free from http://www.r-project.org/. Many students prefer to use the interface RStudio, available for free at http://www.rstudio.com. However, on occassions I will also use Python. Students are free to use any software they prefer. Both R and Python has libraries that implements many network analysis tools.

Grading: The following is a breakdown of the final course grade:

Project proposal: 25%
Final written report: 75%

In judging the report, for reviews, the primary consideration will be depth of understanding reflected in the review (approx 70%), and the remaining 30% grades will be on comprehensiveness of the review. For research project, 50% points will be awarded for displaying a thorough understanding of the problem, and the remaining 50% will be awarded for the progress or potential for progress shown.
The traditional rubric will be used to compute the final letter grade: A: 93 – 100, A-: 90 – 92.9, B+: 87– 89.9: B: 83 – 86.9, B-: 80-82.9, C+: 77-79.9, C: 73 – 76.9, C-: 70–72.9, D+: 67 – 69.9, D: 60–66.9, E: below 60.

***Optional* References:** There is no required textbook for the course. The following are optional references.

- Network analysis Texts
  - Kolaczyk, Eric D., and Gábor Csárdi. *Statistical analysis of network data with R.* New York: Springer, 2014.

  - Newman, Mark. *Networks.* Oxford university press, 2018.
  - Frieze, Alan, and Michał Karoński. *Introduction to random graphs.* Cambridge University Press, 2016.

- Review Articles (More to be added)
  - Goldenberg, Anna, et al. "A survey of statistical network models." Foundations and Trends in Machine Learning 2.2 (2010): 129-233.
  - Matias, Catherine, and Stéphane Robin. "Modeling heterogeneity in random graphs through latent space models: a selective review." ESAIM: Proceedings and Surveys 47 (2014): 55-74.
  - Gao, Chao, and Zongming Ma. "Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing." arXiv preprint arXiv:1811.06055 (2018).
  - Smith, Anna L., Dena M. Asta, and Catherine A. Calder. "The geometry of continuous latent space models for network data." Statistical Science 34.3 (2019): 428-453.

Special Considerations: If a situation exists or arises that you think may hinder your progress in this class, you must notify me as soon as possible.

Academic Misconduct: It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

Disability Services: Students with disabilities that have been certified by the Office for Disability Services will be appropriately accommodated and should inform the instructor as soon as possible of their needs. The Office for Disability Services is located in 098 Baker Hall, 113 W 12th Avenue; telephone 292-3307; http://www.ods.ohio-state.edu/.