UCLA Department of Statistics

STATS 218

Statistical Analysis of Networks

Course Information

Professor:	Mark S. Handcock
Lectures:	Tuesday and Thursday, 2:00pm - 3:15pm, Boelter Hall 5149
Office Hours:	Friday 1:00pm to 2:00pm. In-person and zoom at this link Also feel free to post a question on the Bruin Learn Discussion board any time.
Office Location:	8105C Mathematical Sciences Bldg or zoom at this link

Course description: Lecture, three hours. Limited to graduate students. Introduction to analysis of social structure, conceived in terms of social relationships. Major concepts of social network theory and mathematical representation of social concepts such as role and position. Use of graphical representations of network information. S/U or letter grading.

Discussion

We will use the Bruin Learn Discussion board for Q&A, discussions on topics, FAQ, etc. Please use it freely.

Motivation and Synopsis

This course is an introduction to the analysis of social structure, conceived in terms of social relationships.

Social structure is conceptualized as a system of social relations tieing distinct social entities to one another. Social network theory is the attempt to represent the structure in social relations via networks. It is then a theory pertaining to types of observable social spaces and their relation to individual and group behavior.

Observations on the social relations are of two forms:

- individual level information on the social entities
- relational data on pairs of entities

While both forms are important for the study of social relations, social network theory recognizes fundamental role of the relational information. It is based on the premise that social context is an important determinant of individual behavior. It seeks to understand individual and group behavior in terms of relational information rather than as solely the aggregation of individual characteristics.

The focus of the course are modern methods for the statistical analysis of social networks. The course covers the major concepts of social network theory and the mathematical representation of social concepts such as "role" and "position".

Visualization plays a central in social network analysis. With the development of high speed computing and graphical display tools, visualization has become a flexible and powerful tool in the exploration of social relations. Graphics exploit the power of our visual senses to convey information in a direct way. In this course we will emphasize the use of graphical representations of network information as much as possible.

The course will also involve the practical application of the ideas of statistical computing and their implementation through statistical software, particularly R. As statistical computation is essential for many of the modeling approaches, expertise will need to be developed.

Prerequisites

- At least two courses in statistics, one of which is on modeling.
- People may still take this class as long as they have background on matrix algebra, probability theory, and programming skills. To do this attend the first classes and we can assess if this is advisable.

Course Materials

Required Textbook

[**H**] Eric Kolaczyk

Statistical Analysis of Network Data (2009).
Springer

Required and available online as a ebook from the library, free for UCLA students.
http://link.springer.com/book/10.1007%2F978-0-387-88146-1

[**KCR**] Eric Kolaczyk and Gabor Csardi

Statistical Analysis of Network Data with R (2014).
Springer

Required and available online as a ebook from the library, free for UCLA students.
http://link.springer.com/book/10.1007%2F978-1-4939-0983-4

[**KCR**] David Easley and Jon Kleinberg

"Networks, Crowds, and Markets: Reasoning about a Highly Connected World (2010).

Available free online.
https://www.cs.cornell.edu/home/kleinber/networks-book/

Students must be connected to the UCLA network to obtain their free download. Students who would like to download the textbook off-campus may do so by connecting the the UCLA network via VPN https://www.bol.ucla.edu/services/vpn/all.html.

Syllabus of the Course

The syllabus of the course will develop on the following weekly schedule. The some later topics may not be reached and we will make choice among them toward the end of the quarter.

- R network data types, sna and network packages reading and manipulating networks, plotting
- degree, connectedness, cycles, centrality, betweenness, etc. Computing descriptive statistics

Week	Contents
1	Introduction • Social science is the study of relationships and relationships can be represented via social networks • Where we are going
2	Relational phenomena • types of relational data and networks
3	Mathematical foundations of networks • Using graphs to represent social relations • Graph concepts and notation
4	Fundamental statistics of graphs • density, sociality • models for degree distributions
5	Fundamental statistics of graphs • connectivity
6	Fundamental statistics of graphs • centrality and centralization
7	Simple Stochastic Models for Networks • hypothesis and significance testing for graph structure
8	Stochastic Models for Networks
9	Fitting Social Network Models to Data
10	Modeling Cohesive subgroups
11	Introduction to Exponential-family Random Graph Models (ERGMs)
12	More on ERGM • more complex models

	• mode degeneracy
13	Beyond ERGM • Hierarchical ERGM • Exponential-family Random network models
14	Inference and Simulation within the general ERGM framework
15	Social Balance and Transitivity
16	Goodness of Fit of Network Models
17	Latent Position Models for Networks
18	More on Latent Position Models
19	Social Relations Models for Networks
20	Modeling networks when data is missing or sampled
21	Modeling Longitudinal Networks

Learning Outcomes

- Students will explain the concept of subjective probability.
- Students will be able to formulate simple stochastic models for social phenomena.
- Students will be able to explain Bayesian inferential procedures for stochastic models for social phenomena.
- Students will explain the difference between Bayesian and Frequentist statistical analyses.
- Students will be able to write code in Stan and R to implement Bayesian inference.

Course Webpage and Discussion Forum

The course has a webpage through the UCLA Bruin Learn system. The webpage will be continuously updated throughout the course with handouts, homework assignments and solutions. Users sign in to Bruin Learn with their UCLA Logon IDs.

I will be using Bruin Learn Discussion board to provide discussion of issues in class and related questions. For questions that might be of interest to other students, please use it to ask a question rather than solely emailing me. There other students and the TA can answer questions in addition to me. Example of questions are about interesting articles you have seen in the media, problems with access to resources, homework or computer questions. Enjoy!

Please regularly consult this classes Bruin Learn home page, the Discussion board and the archive of the Announcements mailing list. It will contain lecture notes, homework, solutions and course information.

Computer Usage and Software

The computer is the scientific laboratory of the applied researcher in quantitative fields. As such this course requires the student to develop a degree of comfort and competence “in the lab”.

Our computer interface to R will be the RStudio IDE, which you can download from Posit.

Course Requirements and Grades

- 50% Homework (4-5 assignments, weighted equally, none are dropped)
- 50% Final project (written report)

Homework

There will be regular homeworks on the theory, modeling and data analysis. All homework assignments will be made available from the Bruin Learn *Assignments* page. Students will submit the solution as PDF files electronically via the gradescope link on the Bruin Learn home page.

Students will be graded on a scale of 1 to 10 for each homework. None of the homework scores will be dropped. It is your responsibility to verify that your homework assignment successfully uploaded by the deadline.

Files must not refer to any resources on the local machine or to files that are not publicly available online. No one should make manual edits to a data file on his or her local machine.

Students are free to discuss homework problems and solutions. Discussing the contents of the course with fellow students can be a valuable element of the learning process, and doing so is therefore generally encouraged. However, each student must hand in their own solutions, and the student should, if asked, be able to explain the solutions.

The project can take one of two forms. You should choose one only.

The **first** is to review a topic in statistical modeling of networks.
This topic should extend or expand on a topic covered in the class, but not be just that topic. The report should be stand-alone, starting from the level of the lecture notes and expanding out to review the existing literature on that topic. It should be written at the level so that a fellow student can read it (after they have taken the course). It need not analyze real data, but may use real or simulated data to illustrate the main points. The length can vary, but projects from prior times I have taught this class have been 15 pages in length. If you choose this form you must pass the topic by me first by sending me a description in an email, chatting with me after class or coming to my office hours.

The **second** is to undertake an an analysis of a social network that you find interesting. You can select any network data-set you find interesting, but preferable related to your graduate work. I do not want a quick and routine analysis; a good job will show understanding of the problem and possible solutions and techniques to consider. The technical results should be stated clearly. The report must contain a clearly written conclusion section giving a non-technical summary that is concise and informative. The data set should contain at least 20 nodes, and at least two variables measured for each node. Do not merely use data from a textbook - the world is an interesting place! All data sources must be cited, and described.

For details of the project, see the *Assignments* section.

Late Submissions

There is a 10 minute grace period. Submissions up to 10 minutes late will be accepted with no penalty.

Homework assignments submitted 11 minutes late or more will be accepted with penalty. There is a minimum deduction of 1 point for being less than one hour late. An additional 0.5-point deduction will be taken for each additional hour it is late.

Thus, an assignment that is between 11 and 59 minutes late will receive a 1-point deduction. An assignment that is between 1 hour and 1:59 late will receive a 1.5-point deduction. An assignment that is between 2 hours and 2:59 late will receive a 2-point deduction, and so on.

I strongly advise uploading the homework to the Bruin Learn well in advance of the deadline in case there are connectivity problems or server issues.

Comments of suggestions about the course

I welcome comments or suggestions about the course at any time, either in person, by letter, or by email. Please feel free to use these ways make comments to me about any aspect of the course.

University Policies

Academic Integrity

UCLA is a community of scholars. In this community, all members including faculty, staff and students alike are responsible for maintaining standards of academic honesty. As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. You are evaluated on your own merits. Cheating, plagiarism, collaborative work, multiple submissions without the permission of the professor, or other kinds of academic dishonesty are considered unacceptable behavior and will result in formal disciplinary proceedings usually resulting in suspension or dismissal. See the Dean of Students website for more information.

[source: Dean of Students syllabus statement (syllabus)]

Accommodations for Students with Disabilities:

If you are already registered with the Center for Accessible Education (CAE), please request your Letter of Accommodation in the Student Portal. If you are seeking registration with the CAE, please submit your request for accommodations via the CAE website. Students with disabilities requiring academic accommodations should submit their request for accommodations as soon as possible, as it may take up to two weeks to review the request. For more information, please visit the CAE website , visit the CAE at A255 Murphy Hall, or contact us by phone at (310) 825-1501.

[source: Center for Accessible Education (Faculty Questions)]

Resources for Students

UCLA provides resources if you are feeling overwhelmed and need personal and/or academic assistance.

Please see the Red Folder REV2020 web for more information.

Title IX

Advocacy and Confidential Services

Please note that Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 205 Covel Commons, Los Angeles, CA, 90095, care@careprogram.ucla.edu, (310) 206-246 5. Counseling and Psychological Services (CAPS) provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768.

Reporting and Non-confidential Services

Your professor is required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator should he become aware that you or any other student has experienced sexual violence or sexual harassment. In addition, You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2255 Murphy Hall, titleix@equity.ucla.edu , (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491.