36-780: Social Network Models
Spring 2014 (Mini 3)

TTh 10:00-11:50am, BH 235A

http://www.stat.cmu.edu/~brian/763

Course Information

Instructor:	*TA:*
Brian Junker, Statistics	Maricio Sadinle Garcia-Ruiz
132E Baker Hall	Office Wean 8115
268-8874	Phone 268-5610
brian@stat.cmu.edu	msadinle@stat.cmu.edu

Office Hours (BH 132E):	*Office Hours:*
Noon on Tuesdays, 1:30pm on Thursdays	TBA
(or by appointment).	*(or by appointment).*

Prerequisites

There are no formal prerequisites for this class. However, I expect you to be familiar with statistical theory and the statistics of applied linear regression at a senior undergraduate "stat major" or beginning graduate level. As a guide, at CMU I would expect you to know roughly the content of 36-226 or 36-705 (statistical theory) and 36-401 or 36-707 (modern applied regression).

You will also be expected to know, or learn quickly, the computational software used for this course. Primarily we will be using R and packages installed in R. It is also a good idea to know how to use the technical typsestting system LaTeX.

Please feel free to contact me if you have any questions or need additional information.

Required Text

There is no required text for this course. I will make available handouts, class notes, research papers from the academic literature, etc., as needed. You will also need to use the help facility in R, Google, Google Scholar, etc., to find some information for the course, on your own.

Other Worthwhile Texts (NOT Available in the Bookstore)

Two interesting but somewhat out-of-date textbooks are:

- de Nooy, W., Mrvar, A., & Batagelj, V. (Eds.). (2005). *Exploratory social network analysis with Pajek (Vol. 27).* Cambridge University Press.
- Kolaczyk, E. D. (2009). *Statistical analysis of network data.* Springer.

We will be reading some survey articles covering more recent developments. One, almost book-length, is

- Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning, 2(2),* 129–233.

Most of the newest material that we will dicuss is described briefly at

- http://hnm.stat.cmu.edu

Course Description and Course Objectives

"Everybody knows" about online social networks such a Facebook, Google+, LinkedIn, etc., but in fact, as Fienberg (2012) says, *"Many scientific fields, especially in the social sciences, have used network representations for their problems. The 'statistically' oriented literature on the analysis of networks derives from a handful of seminal contributions, such as those by Simmel and Wolff (1950) at the turn of the last century, and by Moreno (1934), who in the 1930s invented the sociogram—a diagram of points and lines used to represent relations among persons, a precursor to the graph representation for networks."*

In this course we will briefly review the descriptive analysis of social network data that dominated the field until quite recently. We will take a closer look at several successful generative statistical models for social networks, and in particular we will consider the *conditionally independent dyad (CID) models* and *hierarchical network models (HNM's)* described at `http://hnm.stat.cmu.edu`. A key goal will be to try out the new `CIDnetworks` software for R, being developed by Andrew Thomas and colleages, for these models. In addition we will read and present in class literature on social network analysis that interests us, and complete short projects.

Thus, my goals for you in this class are:

- To understand at a high level the descriptive analysis of social network data.
- To understand how a generative model entails a statistical model, and to understand the advantages of statistical models over descriptives, in offering an avenue for combining analyses across ensembles of networks, or extending analyses from smaller samples to larger ones, etc.
- To engage some current research questions in social network analysis.
- To apply what you have learned to a small project of your own.

Since the course is only 7–8 weeks long, we can do nothing in detail. The idea is to give you a taste of each topic, so that when you need to do this kind of analysis in the future, you will have some idea what details are important, and where to look to find out about them.

Computing

Our work in this course will depend primarily on R, and packages for doing social network analysis in R. Some familiarity with procedural and object-oriented programming (in C, Java, Python, R, etc.), and with manipulating and traversing data structures, will be very helpful. There are no commonly-held standards for organizing social network data for analysis, and many different standards exist side-by-side in various R packages.

The first couple of assignments will serve as both an introduction to descriptive analysis of social networks, and as a "crash course" in R, but if you do not already know R reasonably well, you will have to do some additional work on your own.

You are also welcome to use other software tools, but you must document your work carefully in that case, so that we can follow what you are doing.

Tentative Schedule

Week	Date	Tentative Topics	Tentative Work
Week 1	Jan 14, 16	Course Introduction, Descriptive Network Analysis	Stanford Labs 1, 2, 3; Fienberg (2012)
Week 2	Jan 21, 23	ERGM's, Latent Space Models, Generative Statistical Models, CID Models	Stanford Labs 7 (partial) and 8
Week 3	Jan 28, 30	From Multiple Blocks to Multiple Networks: (MM)SBM's and Hierarchical Network Models	Papers, working with CIDnetworks & related software
Week 4	Feb 4, 6	Special Topics or Guest Lectures	Goldenberg et al. (2009). Scheduling paper and project presentations
Week 5	Feb 11, 13	Special Topics or Guest Lectures	TBA; Scheduling paper and project presentations
Week 6	Feb 18, 20	Paper and project presentations	TBA
Week 7	Feb 25, 27	Paper and project presentations	TBA
Week 8	Mar 4, 6	Paper and project presentations	TBA

Student Work

Your work in the course will consist of

- Computer "labs" and other assignments
- Reading papers for the class
- Posting questions (and answers!) in the Blackboard discussion area for this class, and participating in live class discussions
- Projects or paper presentations.

Very roughly, your grade will be determinded in these proportions:

Material	Proportion of Grade
Submitted Assignments	10–20%
Blackboard posts & class participation	30%
Project or paper presentation	50–60%

Anything to be submitted for grading should be prepared as a *single* pdf document, and submitted electronically. There will be more instructions about submitting written work as the course progresses.

Please label all output, plots, variables, etc., appropriately. Always be judicious about including computer output and graphs: show enough that we can clearly see what you are doing, but not so much that we will get lost or bored leafing through your work! A good rule of thumb is to remove any figures, tables, graphs, etc. that you do not have something interesting to say about in the text of your homework solution.

Projects & Paper Presentations

Each registered student in the class will be required to do a small project, or report on a published paper or papers for the class.

- If you have a project on social network modeling that is a part of your research, I encourage you to make that be your project for the mini.
- If you do not have a network data analysis or modeling project, then you should select one or more research papers to present to the class.

In both cases you will need to

- Lead a class discussion on your paper(s) or project, using projected slides or other tools as appropriate.
- Write & submit a short "conference paper" on your work.

More details on this will be forthcoming as the mini gets underway.

Paper presentations and projects will be evaluated on the following points:

- Is the content of the paper or project interesting and worth knowing about?
- Does it take some effort to understand the material?
- Did you do a good job in making the material comprehensible to listeners/readers in your class discussion and short written report?

Academic Integrity

As members of a top-ranked academic institution, your academic integrity is assumed and expected.

For all work, if you get ideas or words from a website, journal article, book, another person (in or out of this class), etc., cite the source in your writeup, right where you use it. Then put a bibliography or list of sources cited at the end of the writeup.

Carnegie Mellon guidelines are listed at http://www.cmu.edu/academic-integrity/; however, I expect each of you to behave well above these lower bounds.

Disability and other Special Needs

Carnegie Mellon makes great efforts to provide physical and programmatic campus access to everyone. Disability Resources ensures that qualified individuals receive reasonable accommodations and that they further receive the rights and protections to equal access programs and services as guaranteed by the Americans With Disabilities Act (ADA) and Section 504 of the Rehabilitation Act of 1973.

If you have a documented disability, please let me know so that we can take whatever steps are needed to accomodate your needs.

Please contact CMU's Disability Resources office (http://www.cmu.edu/hr/eos/disability/) if

- You think you may have a disability and want to document it;
- You have a documented disability that is not being adequately accomodated.

For other issues and special needs, please contact me, your advisor or another trusted mentor, and/or the Office of the Dean of Student Affairs (http://www.studentaffairs.cmu.edu/dean/).