Statistical Analysis of Network Data

Arizona State University
Criminology and Criminal Justice (CRJ) 605
Spring Semester 2024

Time: Wednesdays, 1:30-4:15 PM
Room: Dtphx CRONK 124

Professor: Jacob Young, Ph.D.
Office Hours: Tuesdays 1:30-2:30 (please reserve at: https://calendly.com/jacobtnyoung/crj-605-office-hours)
Office: UCENT Suite 649
Email: jacob.young.1@asu.edu
Course Website: https://jacobtnyoung.github.io/SAND
Course Issues/Help Website: https://github.com/jacobtnyoung/SAND/issues
Course Canvas Website: https://canvas.asu.edu/courses/173880

Overview

Network *science* is a paradigm (i.e. a model of how the world works) that takes as its domain of interest the interdependence among units. This paradigm examines patterns or regularities in relationships (i.e. *structure*) among interacting units and focuses on a) how such patterns influence the behavior of these units and b) how such patterns are generated. Network *analysis* characterizes a broad class of techniques for describing and making inferences about research questions generated within this paradigm.

This course provides an introduction to these techniques for analyzing network data. The overarching goal of the course is that, upon completion, you will be capable of developing research questions from a network perspective and incorporating network-based tools in your own research. The course is designed to be primarily methodological, dedicating the majority of our time to working through the mechanics of network-based tools. You will gain experience with R and RStudio, a commonly used software program for network data management, analysis, and visualization, and you will learn to use data-driven documents to create reproducible workflows.

The course is divided into two primary sections. First, we will examine what network data "look like," how to work with these data in R, and how to describe the properties of a network (i.e. descriptive statistics). The second section of the course will focus on inferential analysis of networks (i.e. hypothesis testing on cross-sectional and longitudinal data structures).

Prerequisites

Graduate level probability and statistics (including standard hypothesis testing and regression methods) is assumed. Additional mathematical and computational background is not required. Prior knowledge of network analysis is not assumed. Prior experience with R and RStudio is not assumed.

Course Materials

Required Texts:
There are 2 books required for this course:

Wasserman, Stanley and Katherine Faust. 1997. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
Luke, Douglas A. 2015. *A User's Guide to Network Analysis In R.* New York: Springer.

Additional readings for each week are available through links embedded in this document.

Online Resources, Listservs and Websites:
I would encourage you to visit https://jacobtnyoung.github.io/sna-textbook, an online social network analysis textbook I have assembled.

I recommend signing up for two listservs for this course: SOCNET (https://lists.ufl.edu/cgi-bin/wa?SUBED1=SOCNET&A=1) is a general social network listserv for posting questions. I like this listerv because you get a sense of what kinds of questions people are asking in network science. Rbloggers (https://www.r-bloggers.com/) is a daily email that covers a variety of topics in R. I like this listserv because it provides a robust coverage of what can be done in R, what people are doing with R, and always provides the code for reproducing what is being discussed.

You should also take a look at the International Social Network Analysis (INSNA, https://www.insna.org/) website. This is the main professional association for social network researchers. If you really want to get involved with network analysis, go to INSNA's annual conference, *Sunbelt* (see INSNA website for details). At a minimum, you should consider reviewing articles in *Social Networks* and *Network Science*.

Software:
There are many software options for network analysis and visualization. In this course, we will rely exclusively on R. R is a very flexible tool for a wide range of statistical analyses. If you are not familiar with R do not worry, we will be spending a lot of time working through examples. R can be downloaded at: http://www.cran.r-project.org.

We will primarily work with R in RStudio. It is a free, open-source IDE (integrated development environment) for working with R. RStudio can be downloaded from: http://www.rstudio.com.

Want a non-iatrogenic "boot-camp" experience with R? Materials for an R workshop I teach in the summer are available on a page at my website: https://jacobtnyoung.github.io/RWorkshop. I would encourage you to work through the labs as they are thoroughly commented and are very basic.

NOTE: I would strongly encourage you to bring a laptop to every class. If necessary you can share with one other person. Please try to keep computer sharing to dyads (i.e. avoid triads).

Github Issues:
We will be using GitHub Issues for posting questions about this course. The GitHub Issues page for this course is: https://github.com/jacobtnyoung/SAND/issues. Github Issues is a powerful tool for addressing problems you are encountering as you work with the materials for this course. You will need to create a Github account to use the page. You can do that here: https://github.com/. I will go over this during class time.

Course Requirements

Your grade in this course will be based on homework assignments (30% of your grade) and the preparation and presentation of an original research paper based on the course content (70% of your grade).

1. Homework: Roughly each week you will be asked to complete a homework assignment based on the week's content. The goal of the assignments is to familiarize you with the concepts **and** the software. You will **really** learn how to "do" network analysis by getting your hands dirty using the software and working with real data. Working on homework with your classmates is acceptable and a good way to learn. However, students must complete their **own** analyses and submit their own write-ups of the analyses. For each homework I will provide a template for you to use for the assignment. Homework are graded based upon an assessment of whether you have sincerely attempted the assignment and answered over half of the questions correctly. This is designed to hold you accountable for the material, but not create anxiety about perfection. I will post a set of solutions to the course website the day the homework is due and we will work through the homework in class that day. There are six (6) homework assignments and each assignment is worth 50 points for a total of 300 points.

2. Course Project: Each student will prepare an original research paper using either network data or a network concept applied to the topic of the student's choice. Topics will be chosen by you **and approved by me**. As with any paper you write in a graduate seminar, your aim should be the development of a publishable manuscript. Students are **encouraged** to work in groups, but groups can be no larger than two students.

To help you make progress on the project, I require you to hand in 2 brief *research proposals* (these proposals are ungraded and are simply for your benefit). The first proposal (a paragraph is fine) should describe generally what your topic is, what questions you are asking, and what data you might use. The second proposal is more detailed and should include your research questions, hypotheses (if appropriate), and identify relevant literature and theory, describe the dataset, and indicate analyses you will perform. Your project may evolve after the second proposal is submitted, but I want you to begin thinking about all of these issues early. Also, by this time, you should have the data you are going to use. The second proposal should be 2-3 pages. On the same day you submit your second proposal, you will give a *mini presentation* discussing the content of your second proposal. The presentation is graded and is worth 100 points. Rubric for grading the presentation will be provided prior to the mini presentation to help you prepare. Due dates for each proposal and mini presentation are given in the course calendar.

During the last class session, students will present their data/findings to the class in the format of a conference style *presentation*. The full presentation is graded and is worth 200 points. Rubric for grading the presentation will be provided prior to the full presentation to help you prepare.

The *final paper* is to be no longer than 25 pages in length, but no less than 2,500 words double-spaced, 1" margins, and 12-point Times New Roman font. The written paper is due by 5 pm on **May 1st**. The final paper is worth 400 points. Rubric for grading the final paper will be provided prior to due date to help you prepare.

In all, your grade breakdown will be as follows:

Homeworks (50 points x 6)	300 points
Mini Presentation	100 points
Full Presentation	200 points
Final Paper	400 points
Total	1000 points

Your grade will be assigned based on the following scale (passing grades are in **bold**):

A+ 980-1000	**A 920-979**	**A- 900-919**
B+ 850-899	**B 800-849**	B- 750-799
C+ 700-749	C 600-690	D 0-599

Keys to Success

Time Investment: The Arizona Board of Regents, the governing board for ASU, has a policy for how much time students should invest in their courses: "A minimum of 45 hours of work by each student is required for each unit of credit" (http://azregents.asu.edu/rrc/Policy%20Manual/2-224-Academic%20Credit.pdf). As a rule of thumb, for every hour of class you should expect to dedicate 3 hours outside of class to the course. This is the investment needed to succeed in this course.

Time Budgeting: For each week's workload, I strongly encourage you to create a time budget (just Google [or Ask Jeeves] "time budget examples"). Budgeting your time will help you get through the reading and provide sufficient time to work with the analytic components of the course.

Work with Others: Your fellow classmates are an excellent resource for developing your understanding of the material. I would encourage you to find a study partner or form a study group. Among other things, working with others helps create socially enforceable deadlines (i.e. accountability). Also, syntax is a quasi-public good with the property of non-subtractability!

Care about your Work: I understand that you have competing time commitments, that there are multiple demands on your day, and that you may come to this particular course with varying levels of interest. Not everyone will make network analysis a part of their career, but a primary concern I have is that, independent of your long-term substantive interests, you take pride in all of the work that you do in this course.

Miscellaneous Course Information

Late Assignments: I will not accept late assignments and they cannot be made up.

Extra Credit: There is no extra credit in this course.

Honesty: Information regarding cheating and plagiarism can be found at: http://provost.asu.edu/academicintegrity. I strongly encourage you to read this not only for my course, but to protect your interests as a student.

Students with special needs: If you require special accommodations for class, please see https://eoss.asu.edu/drc for ASU policies and procedures. The Disability Resource Center will provide you with a letter noting the classroom modifications that you will need to fully take part in class activities. With this letter, contact me during my office hours and I will gladly make the appropriate adjustments.

Course Evaluation: Teaching evaluations are very important and I strongly encourage you to take the time to complete it. Your participation is essential for improving the course. Evaluations are easy to access, just go to MyASU, click on My Classes, and then click Course Evaluations. Select CRJ 605 and you are ready to go. I will remind you when the evaluation survey for this course is available.

COURSE OUTLINE

1/10: Introduction to Network Analysis

Assignments: Homework #1 distributed.

Required Readings:

Network Concept(s)

Wasserman and Faust, Chapter 1: Social Network Analysis in the Social and Behavioral Sciences.

Wasserman and Faust, Chapter 2: Social Network Data.

Criminological Application

Papachristos, A. V. (2014). The Network Structure of Crime. *Sociology Compass, 8*(4), 347-357.

1/17: Introduction to R/RStudio, Data-Driven Documents, and Github Issues

Required Readings:

Network Concept(s)

Luke, Chapter 1: Introducing Network Analysis in R.

1/24: Network Data Structures & Network Data in R

Assignments: Homework #1 *DUE*, Homework #2 distributed, create a Github account (https://github.com/).

Required Readings:

Network Concept(s)

Wasserman and Faust, Chapter 3: Notation for Social Network Data.

Wasserman and Faust, Chapter 4: Graphs and Matrices.

1/31: Network Description & Visualization

Assignments: Brief Proposal #1 *DUE*.

Required Readings:

Network Concept(s)

Luke, Chapter 2: The Network Analysis 'Five-Number Summary'.

Luke, Chapter 3: Network Data Management in R.

Luke, Chapter 4: Basic Network Plotting and Layout.

Luke, Chapter 5: Effective Network Graphic Design.

2/7: Review Homework #2 & begin Network Position: Degree Centrality
Assignments: Homework #2 *DUE*, Homework #3 distributed.

Required Readings:
Network Concept(s)
Luke, Chapter 7: Actor Prominence.
Wasserman and Faust, Chapter 5: Centrality and Prestige.

Criminological Application
Thomas, D. R., & Wahedi, L. A. (2023). Disrupting hate: The effect of deplatforming hate organizations on their online audience. *Proc Natl Acad Sci U S A*, *120*(24), e2214080120.

2/14: (MORE) Network Position: Closeness and Betweenness Centrality
Assignments: Mini-Presentation instructions distributed.

Required Readings:
Criminological Application
Faris, R., & Felmlee, D. (2011). Status Struggles: Network Centrality and Gender Segregation in Same- and Cross-Gender Aggression. *American Sociological Review*, *76*(1), 48-73.

2/21: Review Homework #3 and begin Two-Mode Networks/Bipartite Graphs
Assignments: Homework #3 *DUE*.

Required Readings:
Network Concept(s)
Luke, Chapter 9: Affiliation Networks.
Wasserman and Faust, Chapter 8: Affiliation and Overlapping Subgroups.

Criminological Application
Young, J. T. N., & Ready, J. T. (2014). Diffusion of Ideas and Technology: The Role of Networks in Influencing the Endorsement and Use of On-Officer Video Cameras. *Journal of Contemporary Criminal Justice, 31*(3), 243-261.

2/28: Mini-Presentations; Projection & Weighted Graphs
Assignments: Brief Proposal #2 *DUE*, Homework #4 distributed.

3/6: No Class (*Spring Break*)

3/13: Review Homework #4; Cross-Sectional Network Analysis using Exponential Random Graph Models
Assignments: Homework #4 *DUE*, Homework #5 distributed.

Required Readings:
Network Concept(s)
Luke, Chapter 11: Statistical Network Models.

Criminological Application
Young, J. T. N., & Haynie, D. L. (2020). Trusting the Untrustworthy: The Social Organization of Trust Among Incarcerated Women. *Justice Quarterly*, 553-584.

3/20: (MORE) Cross-Sectional Network Analysis using Exponential Random Graph Models

3/27: Review Homework #5 & Longitudinal Network Analysis using Stochastic Actor-Based Models
Assignments: Homework #5 *DUE*, Homework #6 distributed.

Required Readings:
Network Concept(s)
Luke, Chapter 12: Dynamic Network Models.

Criminological Application
Turanovic, J. J., & Young, J. T. N. (2016). Violent Offending and Victimization in Adolescence: Social Network Mechanisms and Homophily. *Criminology*, *54*(3), 487-519.

4/3: (MORE) Longitudinal Network Analysis using Stochastic Actor-Based Models
Required Readings:
Criminological Application
Kreager, D. A., Schaefer, D. R., Davidson, K. M., Zajac, G., Haynie, D. L., & De Leon, G. (2019). Evaluating peer-influence processes in a prison-based therapeutic community: a dynamic network approach. *Drug Alcohol Depend*, *203*, 13-18.

4/10: Review Homework #6 & Dynamic Visualization of Networks *or* Sequential Temporal Exponential Random Graph Models and Epidemic Modeling
Assignments: Homework #6 *DUE,* Presentation instructions distributed.

4/17: Guest Lecture (Marva Goodson-Miller): Ego-Centric Social Networks
Assignments: Final paper instructions distributed.

Required Readings:
Criminological Application
Goodson-Miller, M. V. (2022). A First Look at Justice-Involved Women's Egocentric Social Networks. *Social Networks*, 70, 152-165.
Marsden, P. V., & Hollstein, B. (2023). Advances and innovations in methods for collecting egocentric network data. *Soc Sci Res*, *109*, 102816.

4/24: Presentations & Conclusion

5/1: Final Paper Due

BRIEF **COURSE OUTLINE**

1/10: Introduction to Network Analysis

1/17: Introduction to R/RStudio, Data-Driven Documents, and Github Issues

1/24: Network Data Structures & Network Data in R
Assignments: Homework #1 *DUE*, create a Github account (https://github.com/).

1/31: Network Description & Visualization
Assignments: Brief Proposal #1 *DUE*.

2/7: Review Homework #2 & begin Network Position: Degree Centrality
Assignments: Homework #2 *DUE*.

2/14: (MORE) Network Position: Closeness and Betweenness Centrality

2/21: Review Homework #3 and begin Two-Mode Networks/Bipartite Graphs
Assignments: Homework #3 *DUE*.

2/28: Mini-Presentations; Projection & Weighted Graphs
Assignments: Brief Proposal #2 *DUE*.

3/6: No Class (*Spring Break*)

3/13: Review Homework #4; Cross-Sectional Network Analysis using Exponential Random Graph Models
Assignments: Homework #4 *DUE*.

3/20: (MORE) Cross-Sectional Network Analysis using Exponential Random Graph Models

3/27: Review Homework #5 & Longitudinal Network Analysis using Stochastic Actor-Based Models
Assignments: Homework #5 *DUE*.

4/3: (MORE) Longitudinal Network Analysis using Stochastic Actor-Based Models

4/10: Review Homework #6 & Dynamic Visualization of Networks *or* Sequential Temporal Exponential Random Graph Models and Epidemic Modeling
Assignments: Homework #6 *DUE*.

4/17: Guest Lecture (Marva Goodson-Miller): Ego-Centric Social Networks

4/24: Presentations & Conclusion

5/1: Final Paper Due