THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

# SYLLABUS: STAT 8810
## STATISTICAL INFERENCE IN NETWORK DATA
## SPRING 2025

# Course overview

## Instructor

Instructor: Subhadeep Paul
Email address: paul.963@osu.edu
Lectures: Tuesdays and Thursdays, 12:40pm–2:30pm in Baker Systems 136, starting February 27 2025.
Office hours: Cockins Hall 231, Tuesdays 4:30 – 5:30 PM.

## 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. We will discuss a variety of methods for analyzing network data and learn to apply those methods to datasets using R packages.

**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 property and heterogeneous degree distribution

- Data structures - single network, multi-layer network, time varying network, multimodal network

**Module 2: Random graph models**

- Erdos-Renyi random graphs and their properties.
- Probabilistic techniques and concentration inequalities for random graphs.
- Theoretical results on properties of random graph models

**Module 3: Stochastic Block Models**

- Stochastic Block Models – definition and properties, estimation methods - maximum likelihood, modularity optimization, variational methods, spectral methods
- Characterizing the error of community detection and evaluation of the methods.
- Consistency and minimax rates of community detection using spectral and likelihood methods.
- Extensions and generalizations - degree corrected, mixed membership, superimposed models, random dot product graphs, graphons.

**Module 4: Latent Space Models**

- Modeling choices – social relations model, additive and multiplicative effects model
- Estimation and Inference
- Extensions - hyperbolic space, more complex data structures

**Module 5: Dynamic and Multilayer networks**

- Multilayer networks: models and methods
- Discrete time dynamic networks
- Continuous time interaction networks

**Module 6: Network peer effects and network experiments**

- Peer effects and homophily
- Network experiments
- Network time series

# Course learning outcomes

Upon successful completion of the course, students will be able to

1. Compute various summary statistics and properties of network data
2. Fit statistical models to network data and assess model fit
3. Theoretically and through simulations evaluate the accuracy of various methods of network analysis
4. Perform statistical learning tasks on network data including community detection, link prediction, and classification
5. Identify appropriate analysis strategies for more complex networked data structures including dynamic and multilayer networks
6. Estimate network peer effects and design network experiments.
7. Apply the network analysis methods to network data obtained from a variety of disciplines.

# Course materials and technologies

## Textbooks

**Suggested**

***Statistical Analysis of Network Data with R (2020)*** by Eric D. Kolaczyk and Gábor Csárdi, Springer New York, NY

The electronic version of this book is freely available to OSU students from the OSU library at this link (please log in with your OSU credentials)

https://library.ohio-state.edu/record=b9007501

(click on “connect to resource SpringerLink” to download a PDF copy).

**Course material will be supplemented with typed lecture notes that will be provided regularly.**

### Necessary software

- **I will use the statistical software package called R** (The R Project for Statistical Computing; http://www.r-project.org/) for illustrating data examples. This software package is available as Free Software.
    - You can download R for Windows, Mac, and Linux, from the CRAN archive at https://cran.r-project.org.
    - An in-depth introduction to R is available at http://cran.r-project.org/doc/manuals/R-intro.pdf
- An easier to use interface to R is available in the software package RStudio (now posit). This package is available for Windows, Mac, and Linux and can be downloaded for free from https://posit.co/. **Note that RStudio requires R to be installed.**
- **For the projects, you may use Python or any other programming language/software platforms.**

# Grading and faculty response

## Grades

| Assignment or category | Percentage |
|---|---|
| **Project proposal** | **25** |
| **Final project presentation** | **25** |
| **Final project written report** | **50** |
| **Total** | **100** |

## Grading scale

93–100: A
90–92.9: A-
87–89.9: B+
83–86.9: B
80–82.9: B-
77–79.9: C+

73–76.9: C
70 –72.9: C-
67 –69.9: D+
60 –66.9: D
Below 60: E

## Assignment information

**Project:** There is no homework or exam for this course. The evaluation is based on a written project proposal and final project presentation and report due at the end of the course (the tentative date of the presentation is given below).

There are three options for a project:

1. **Micro-review:** Write a micro-review of the literature on a topic (approximately 5-10 pages). I will provide a list of suggested topics to write the review on. 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. The topic should be broad enough to be able to write a substantial review on, yet not too broad that it becomes difficult to review in a short writeup.
2. **Short Research Report**: You may also perform original research on a topic related to network data and submit a short research report (minimum 5 pages and no maximum). I can provide a number of research topics that I am interested in. If you take the research route, you need to declare a stated research goal (even if vague) in the proposal document 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. You may use existing (unpublished) research that you may already be conducting related to network data as part of your project, but please get permission from other co-authors or advisors on presenting the research during our project presentation and submitting the short research report.
3. **Network data analysis:** You may also analyze a network dataset using methods learnt in the class to obtain scientific insights. The submitted report should be minimum 5 pages with no maximum. I will provide some sources of network data, but you are free to choose any dataset you like. It is expected that you will choose a data and scientific problem such that the data analysis task will be substantial. You may also use data you have collected or analyzing as part of your research and you will not have to share the data with me. However, as in part (2), if you are using existing research for the project, be sure to take permission from co-authors and advisors for presentation and short written report.

You are welcome to work on a project alone or form short groups (maximum 2 members in a group). I expect the effort and quality of the documents to increase linearly with the number of authors**. For all three options, 2 reports and a presentation are required: a project proposal (approx. 1 page), a final written report (see above for suggested lengths), and a final presentation on the last day of scheduled class**.

**Tentative Dates for exams and presentations:**

**Final project presentation: Thursday, April 17, 2025, during class time (the last scheduled class).**

## Faculty feedback and response time

I am providing the following list to give you an idea of my intended availability throughout the course.

### Grading and feedback

For project proposal, you can expect feedback within **7 days**. The final project presentation and report will be promptly graded.

### E-mail

I will reply to e-mails related to course logistics and technical questions that can be answered over email within **48 hours on school days**. However, if you have questions on understanding the course materials or related to your project, please consider visiting my office hours.

# Other course policies

## Academic integrity policy

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/**.

**If I suspect that a student has committed academic misconduct in this course, I am obligated by university rules to report my suspicions to the Committee on Academic Misconduct.** If COAM determines that you have violated the university’s *Code of Student Conduct* (i.e., committed academic misconduct), the

sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the university.
If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me.
Other sources of information on academic misconduct (integrity) to which you can refer include:

- Committee on Academic Misconduct web page (go.osu.edu/coam)
- *Ten Suggestions for Preserving Academic Integrity (*go.osu.edu/ten-suggestions*)*

## Copyright for instructional materials

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

## Statement on Title IX

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.edu or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu

# Your mental health

As a student you may experience a range of issues that can cause barriers to learn, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting ccs.osu.edu or calling 614-292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available 24/7 by dialing 988 to reach the Suicide and Crisis Lifeline.

# Accessibility accommodations for students with disabilities

## Requesting accommodations

**The university strives to maintain a healthy and accessible environment to support student learning in and out of the classroom. If you anticipate or experience academic barriers based on your disability (including mental health, chronic, or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion.**

**If you are isolating while waiting for a COVID-19 test result, please let me know immediately. Those testing positive for COVID-19 should refer to the Safe and Healthy Buckeyes site for resources. Beyond five days of the required COVID-19 isolation period, I may rely on Student Life Disability Services to establish further reasonable accommodations. You can connect with them at slds@osu.edu; 614-292-3307; or slds.osu.edu.**

## Accessibility of course technology

This course requires use of CarmenCanvas (Ohio State's learning management system) and other communication and multimedia tools. If you need additional services to use these technologies, please request accommodations with your instructor.

- Canvas accessibility (go.osu.edu/canvas-accessibility)
- Streaming audio and video
- CarmenZoom accessibility (go.osu.edu/zoom-accessibility)
- Collaborative course tools

# Course schedule (tentative)

Refer to the Carmen course for up-to-date assignment due dates.

| Class | Date | Topics |
|---|---|---|
| 1 | 02/27 | Descriptive summaries and properties of graphs, visualization of network data |
| 2 | 03/04 | Centrality measures, clustering coefficient, small-world property, Community structure and modularity |
| 3 | 03/06 | Properties of Erdos-Renyi random graphs, simulation, Concentration inequalities and theoretical results |
| | 03/11, 03/13 | Spring break – no classes |
| 4 | 03/18 | The stochastic block model (SBM), maximum likelihood estimation |
| 5 | 03/20 | Variational EM method, fitting SBM to networks |
| 6 | 03/25 | Spectral methods for community detection and SBM, Project proposal due |
| 7 | 03/27 | Accessing accuracy of community detection – theory and simulation-based evidence, Consistency and minimax theory for SBM community detection |
| 8 | 04/01 | Extensions of SBM- degree corrected, mixed membership, random dot product graphs, graphons |
| 9 | 04/03 | Latent space models (LSM) – motivation and description, Estimation and inference in LSMs, fitting LSMs to network data |
| 10 | 04/08 | LSM – additive and multiplicative effects, assessing model fit |
| 11 | 04/10 | Multilayer networks – examples, descriptive properties, models and methods, dynamic network models and methods |
| 12 | 04/15 | Causal peer effects, network experiments, network time series |
| 13 | 04/17 | Project presentation, written report due on 04/21. |