

CDS 292: Network Analysis

Fall 2020

Syllabus Part 1

Syllabus is split into three parts:

- **Part 1: General Course Description**
- **Part 2: Instructors**
- **Part 3: Calendar**

Course Introduction

The purpose of CDS 292 is to introduce the novice data analyst to the basic concepts of networks, as well as the mathematical, statistical, and computational tools that support the application of those concepts. Another goal is to introduce healthy practices of data analysis with networks. Networks have become an essential tool in the world of data analysis, and thus in disciplines as disparate as sociology, biology, arts, and physics, networks show up now with abundant regularity.

The course uses the programming language python, which is one of the most versatile and extensively used computational tools at present. The versatility of python in data analysis is well-established and currently it is a tool no data scientists can do without.

CDS 292 is a true amalgamated course, with mathematical, computational, statistical, and data analytic material. To have a sense about what the course tries to achieve, by the end of the course the successful student should 1) be able to perform data analysis using networks and its various concepts, 2) understand the fundamental properties of each concept learned to apply them correctly, 3) be able to error check code and libraries constructed by others (lots of shared code on the Internet has errors) and finally, 4) know enough to understand how to build a networks algorithm from scratch, error check it, and use it.

A 2016 salary survey of young professionals in the Washington, DC, area clearly shows that graduates with backgrounds in computational data sciences are in high demand. In fact, the survey showed that computational data science is among the fastest growing talent needed.

This course is a Quantitative Reasoning Mason Core course, designed to fulfil the relevant learning outcomes. It does this by focusing on the teaching, use, and interpretation of quantitative analysis to study concrete theoretical and real-world problems. **Quantitative reasoning will be expected of students!**

Advice to Succeed in CDS-292:

The course is a big combination of things. It is designed to represent the incredibly challenging new work environment that has developed around the notion of creative/knowledge economies. As in those environments, the key is to know your specialty, but also to be very flexible when it comes to adding/learning new things. In addition, if you are effective in checking your work very well, it will give you a great competitive advantage.

Therefore this course, which is about networks applied to many real-world problems, requires that one becomes proficient at a number of things. Before getting into the details of those, here are some practical ways to increase your chances to do well:

- **Read the book, slides, notebooks and solutions a few times.** There are many little concepts that keep popping up over and over, and they are scattered so unless you read, they will escape you.
- **Practice a lot of code.** Programming is like playing a musical instrument, which can always be done better if you like it, practice, and have some natural talent. I am sure you have all three, but the one you can control is practice. Make variations on the code, try new things, etc.
- **Dust off your old high-school math.** The course does rely on the use of some algebra, matrices, and even clever arithmetic. This goes with Quantitative Reasoning Core designation. Please, please do not underestimate this. Many students join the class proficient in coding but rusty in math. Data Science relies on people understanding the meaning of quantities and numbers. This class fully embraces that notion and there is an expectation that you will use the class to improve those skills.
- **Once solutions are provided, study them.** It cannot be stressed enough how this is essential prep for exams.
- **Take full advantage of collaborative work.** There are activities designed specifically for you to engage with other students. Make use of them. Peer learning is an important part of learning. **But remember your problem set submissions are only your own!**
- **Don't delay in talking to your instructor if there's a problem.** If you wait, the problem may get worse and your instructors may have less options/time to help you through it.

Textbook: The class is covered in the textbook being developed by Dr. López that is made available to you on Blackboard.

Computer IT Support:

- If you are having difficulties with your GMU network account, or with software on the GMU network, please contact IT support: support@gmu.edu . The ITS Support Center is in Innovation Hall, Room 226 or can be reached at: 703-993-8870.

Learning Outcomes and Quantitative Reasoning Mason Core

General:

1. Students are able to interpret quantitative information (i.e., formulas, graphs, tables, models, and schematics) and draw inferences from them. Students are able to decant information to determine if it has network structure.
2. Given a quantitative problem, students are able to formulate the problem quantitatively and use appropriate arithmetical, algebraic, and/or statistical methods to solve the problem. Matrix algebra is part of the methodology of this course.
3. Students are able to evaluate logical arguments using quantitative network theory reasoning.
4. Students are able to communicate and present quantitative results effectively. This is achieved by a set of continuing in-class activities where students are required to apply network techniques to a data problem, analyse the results, and defend them to the class and instructor.

Specific Learning Lesson Outcomes:

These are stated on the Weekly Lessons on the course Blackboard Website.

Lessons

**** SCHEDULE IS SUBJECT TO CHANGE AT THE INSTRUCTOR'S DISCRETION ****

Class Schedule:

Week 1

-Lesson 0: Navigating CDS 292 Online and Lesson 1: python book bootcamp and Colaboratory

Week 2

- Lesson 2: Building Blocks of Networks

Week 3

- Lesson 3 (part 1): Node Degrees, Link Indicators, Network Formulas, and Adjacency Matrix

Week 4

- Lesson 3 (part 2): Node Degrees, Link Indicators, Network Formulas, and Adjacency Matrix

Week 5

- Lesson 4 (part 1): Creating Networks: An Introduction to Network Models

Week 6

- Lesson 4 (part 2): Creating Networks: An Introduction to Network Models

Week 7

- Lesson 5 (part 1): Histograms of Node Degrees

Week 8

- Midterm week: Live review and midterm day

Week 9

- Lesson 5 (part 2): Histogram of Node Degrees

Week 10

- Lesson 6: Paths in Networks

Week 11

- Lesson 7: Shortest Paths in Networks

Week 12

- Lesson 8 (part 1): Triangles, v-shapes, and clustering

Week 13

- Lesson 8 (part 2): Triangles, v-shapes, and clustering

Week 14

- Lesson 8 (part 3): Triangles, v-shapes, and clustering

Week 15

- Special Topic: Epidemics on Networks

Lesson assignments: for due dates of assignments, please consult **Syllabus Part 3**.

Course Material Availability: Lesson slides, videos, python notebooks, problem set assignments, python notebook assignments, solutions, other class materials, and assignments and exam scores will be posted to the class Blackboard site.

Policies and Procedures

Final Course Numerical Grade / Final Course Letter Grade Correspondence:

(Note: I do not anticipate curving grades. Assigned grades are from earned credit.)

Grades Scored Between	Will Equal
97 % and 100 %	A+
94 % and Less Than 97%	A
90 % and Less Than 94%	A-
87 % and Less Than 90%	B+
84 % and Less Than 87%	B
80 % and Less Than 84%	B-
77 % and Less Than 80%	C+
74 % and Less Than 77%	C
70 % and Less Than 74%	C-
60 % and Less Than 70%	D
0 % and Less Than 60%	F

Grading:

- When grading assignments, while correct answers are important, it is more important how you arrived at those answers. For this we look at your derivations, logic used, algorithms, and good problem solving strategies that you write into your assignments. Some (minimal) credit is given for the answer but most credit is given for the correct work leading to the answer.
- Full credit is given for correct answers provided ALL supporting work is shown in an organized manner showing how the answer was arrived at.
- Semester grades will be computed from the following percentages
 - 45% Assignment grade:
 - 15% Colaboratory exercises
 - 30% for problem sets
 - 25% for the midterm exam
 - 30% for the final exam

Assignments:

- All python Problem Set submissions MUST ORIGINATE ON PAPER, then be scanned, and then be sent. For scanning, the gold standard is a standalone scanner. You may use phone apps to scan if you do not have access to a standalone scanner. Some options:
 - Adobe Scan app
 - Microsoft Office Lens

- **Do not send problem sets pictures.** They are hard to read. Think: if you have trouble reading it from a picture, we have even more trouble as we have to read many such submissions.
- **Send a SINGLE pdf document** containing your solutions.
- **CORRECT INDENTATION OF python CODE IS A CONDITION TO CORRECTNESS.** If you are not sure about your handwriting, use a symbol to mark an indented block. Just a right pointing arrow will be enough.
- **Time extensions** may ONLY be granted by the instructor, must be in writing form, and will give a specific time. **Late assignments sent directly to your TA will NOT be graded.**
- The total available points in each Problem Set can vary week to week.
- The semester Problem Set grade will consist of the averaged sum of the individual Problem Set grades. Specifically,

$$\text{Problem Set Grade} = (\text{Sum of points you obtained}) / (\text{Sum of points over all Sets})$$

- **Problem Sets are generally released with slightly over a week before the due date, generally on Friday of the previous week.**

Colaboratory notebook assignments:

- Every week, there will a set of coding problems intended for you to solve with a *programming partner*.
- **One of you will import the provided notebook to Colaboratory, invite his/her partner and start working.**
- An important part of the assignment is that you organize the way you work. This is to avoid overwriting each other's code. There is a video demonstrating how to do this.
- Once you have finished with the notebook. **MAKE SURE THAT AT THE TOP OF THE NOTEBOOK THERE IS A TEXT CELL WITH BOTH OF YOUR NAMES. Then, invite the teaching assistant (mfgirgis1@gmail.com) to the notebook and he will be able to grade it.**
- **To submit your work, both programming partners should go to the submission page in blackboard, leave a message and press the submit button.** This is just so we keep track of the date you sent your work.
- All Notebooks have equal weight.
- The semester Notebook grade will consist of the averaged sum of the individual Notebook grades. Specifically,

$$\text{Notebook Grade} = (\text{Sum of points you obtained}) / (\text{Sum of points over all Notebook})$$

- **Notebooks are generally released with slightly over a week before the due date, generally on Friday of the previous week.**

Exams:

- The exams will be available on Blackboard with the Respondus Lockdown Browsers.
- Once you open the exam you will have 3 hours to complete it unless you open with less than 3 hours before exam closing. In that case, you will only have the remaining time between when you open the exam and the closing time.

- The midterm will be available to be taken on October 14th from 7AM to 10PM.
- The final will be available to be taken on December 10th from 7AM to 10PM.
- **NOTE ON RESPONDUS:** If you have installed RESPONDUS at NVCC or any other college/school/..., it MUST be UNINSTALLED and REINSTALLED for use at MASON. This is due to the internal handshaking that occurs between MASON's Blackboard and RESPONDUS.
- Students will be REQUIRED to show their GMU ID card PRIOR to beginning either the Midterm or Final exams. (Remember: “No GMU ID - No exam grade - No exception”).

More comments on Grading:

- While correct answers matter, it is more important how you arrived at those answers. Completeness in answers (complete thoughts, complete codes, complete plots) is about ½ credit for the problem.
- Semester grades will be computed as a percentage of points earned divided by total points available in each category.
- There are no extra credit assignments/projects available for CDS-292.

Technology Needed and GMU netlogin:

- Each student will require access to Blackboard from GMU. Make sure you have access to your GMU netlogin.
- Each student will require checking email often and therefore also requires have up to date
- Software access for the class:
 - PDF reader with the capacity to display pages in single page view (for slides), and
 - a computer capable of executing python code, and the networkx and matplotlib python libraries,
 - a computer with a browser to use Blackboard and Respondus browser (needs to be Mason Installed or it doesn't work) with a webcam or a phone with a webcam (consult Lesson 0 for details).
 - Access to a Google account is highly recommended (Colaboratory).

Email Requirements:

- All e-mails **MUST** be from your Mason e-mail account. E-mail from non-Mason e-mail accounts will NOT be responded to.
- All e-mails must have “CDS292 Sec”__ (section) at the beginning of the Subject Line so I may easily identify which class you are in.
- Check your Mason email account frequently ideally twice a day once around lunchtime and another at the end of the day “before you go to bed”. At a minimum you must check email once a day.
- Maintain your e-mail boxes so that messages are not rejected for being over quota. And watch that class and GMU e-mails do not end up in your JUNK/SPAM folders.
- You may forward your Mason e-mail to other accounts but always communicate with me (and your fellow students in case you agree to sharing such details) using Mason e-mail for verification of your identity.
- Students are responsible for the content of university communications sent to their George Mason University email account and are required to activate their account and check it regularly. All communication from the university, college, school, and program will be sent to students solely through their Mason email account.
[See <https://masonlivelogin.gmu.edu/login>].

Collaboration & Plagiarism

- All CDS-292 activities are subject to GMU's Honor Code and IT policies.
- **Collaboration:** Students are encouraged to discuss homework problems with each other. Discussion means each student working the problem may talk with someone else (aka: conversation on the logic or software needed to complete the assignment) but will fully work the problem on their own.
- **Plagiarism** will not be tolerated at any time. Students will be given a zero for any assignment, quiz, or exam where plagiarism is suspected. If plagiarism is suspected a second time for any student, an automatic grade of "F" will be assigned for the course with a report sent to the Dean for further action.
- Collaboration becomes plagiarism when: **A STUDENT COPIES THE WORK OF SOMEONE ELSE, EITHER FROM STUDENTS CURRENTLY TAKING CDS-292, OR WHO HAVE TAKEN CDS-292 IN THE PAST OR BY COPYING THE WORK FROM ANY PUBLISHED OR UNPUBLISHED SOURCE.**
- All problem sets, notebook assignments, and exams, including computer programs and associated outputs, turned in for grading represent the student's own work.
- Students may not discuss exam problems with anyone other than your instructor or the CDS-292 TA.
- **COLLABORATION DURING AN EXAM OR QUIZ IS NOT AUTHORIZED AT ANY TIME AND THUS PROHIBITED AND IS CONSIDERED PLAGIARISM!**
- All plagiarism violations will be reported in writing to the Dean's office.

Prohibited Equipment Use During Exams: Exams are closed book/closed notes. The only allowed equipment is the blackboard site under respondus, paper and pencil. Therefore, other computer equipment, smart phone, calculator or other electronic device use during an exam is prohibited. Individual instructors may announce modifications to this policy which will be specific to their class section only.

Disabilities: If you have a documented learning disability or other condition that may affect academic performance students MUST:

- (1) Have the need for accommodation on file with Office of Disability Services (SUB I, Rm. 4205; 703-993-2474; <http://ods.gmu.edu>);
- (2) Provide your instructor with a copy of the Office of Disability Services accommodation determination prior to receiving any accommodations. The instructor will closely protect this information as private and will not share the information with anyone other than the class assistants unless authorized in writing by the student or the Office of Disability Services.
- (3) The instructor can arrange for alternate testing arrangements for the Midterm and Final Exams with PRIOR notice of need, which is needed AT LEAST 1 week in advance although it is better if even more time is allowed.
- (4) PLEASE NOTE: If you are having ANY difficulties with CDS-292 due to personal limitations, PLEASE discuss them with your instructor. We want to help you succeed in CDS-292 and in your GMU career.