

CSC 299: Sophomore Lab in Applied Computing

“Mining Social Networks”

Fall 2018

Tuesdays: CDM 218

Thursdays: CDM 819

Office: (Loop) CDM 841, (Lincoln Park) 990 W. Fullerton, Rm 3130.

Course discussion on Slack (recommended): <https://csc299fall18.slack.com/>

Email: rburke@cs.depaul.edu

Phone: (Loop) 312-362-5910, (Lincoln Park) 773-325-4915

Office hours: Tuesdays 10:00 am – 11:30 am (Loop);

Tuesdays 3:30 – 4:30 pm (Lincoln Park);

Thursdays 4:30 – 5:00 pm (Loop)

Course web site: <http://d2l.depaul.edu/>

Description

In this course, students investigate a particular application of computing. Students learn tools, methodologies, and formalisms used in a particular computing area, and apply them to develop working systems. Courses stress student initiative in investigating the application context, learning new tools (including languages and APIs), studying algorithms and code examples, and working on projects. Topics will vary by the faculty member's interest and perspective.

In this session of CSC 299, we will focus on the analysis of online social networks. We will access the data stored in these systems using their APIs and we will develop a variety of techniques for analyzing and visualizing such data. We will make use of data analysis tools available in Python including Matplotlib and NetworkX, and the open-source graph visualization tool Gephi.

Learning Objectives

Students will be able to:

- develop hypotheses about a network and test them using data collection and analysis.
- use computational tools to visualize complex networks.
- apply basic network analysis metrics such as connectivity, centrality and embeddedness.
- download, prepare, filter and process network data from a variety of sources.
- describe the difference between different models of network formation including random and preferential attachment and their consequences for network structure.
- describe how local mechanisms of attachment in networks can lead to large-scale phenomena such as assortative mixing and small-world linkages over large populations.
- work effectively in a multi-disciplinary group.

Prerequisite:

CSC 242 or CSC 243

Readings

Golbeck, J. *Analyzing the Social Web*. Morgan Kaufmann, 2013. ISBN: 978-0-12-405531-5.

(Optional) Russell, M. A. *Mining the Social Web*. O'Reilly. 2014. ISBN: 978-1-449-36761-9.

Note: all of the Python code for the Russell book can be obtained from the author's GitHub repository at <https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition>

Other documentation and resources available online.

Tools

Jupyter Notebook

We will use Jupyter Notebook, a technology that turns a web browser into an interactive Python session, and various other Python libraries designed for scientific computing and network analysis. The Anaconda installation on the CDM lab computers includes Jupyter Notebook and many of the other libraries that we will use including networkx, matplotlib and others. Anaconda is available for OS X, Windows and Linux.

Gephi

We will also be using the Gephi network visualization tool. This is a Java-based open source tool available from gephi.org.

Slack

We will be using Slack for class discussion and group communication. Slack has a web front-end. The course site is csc299fall2018.slack.com. I will set up student accounts during the first week of class.

Groups

The class will be organized into groups of three or four. Each group will address a different research question in a particular social network. Each student will be assigned a specialist role: data gathering, data analysis and project coordination/reporting. For example, in a four-person group, there will be two data gathering specialists, one data analysis specialist and one project coordinator. See the “Group structure and responsibilities” document for additional information on these roles.

Groups are expected to meet at least weekly outside of class and more frequently as necessary for important milestones. Most assignments will be submitted as a group. Groups will be responsible for regular reporting about their progress. Although some milestones are more closely related to one specialty or the other, all group members are responsible for ensure that the milestones are completely in a timely manner.

Assignments

There are five different kinds of assignments:

- Lab assignments: These are guided lab exercises where the instructor goes through all of the steps in class. These are due the day after the lab, so that student can complete any steps that they missed.
- Quizzes: See below.
- Group lab assignments: These are similar to the individual lab assignments, but only one submission per group is expected. Group lab assignments take place on specialist lab days when we concentrate on one aspect of the project.
- Group homework assignments: Homework assignments provide an opportunity to use the techniques learned in labs to solve larger problems as a group.
- Project milestones: The project milestones mark progress towards the completion of the final project at the end of the quarter.
- Journal entries: Starting Week 2, a journal entry is due every Friday. The purpose of the journal entries is to help students reflect on their group project work and to document steady progress towards project goals.

Quizzes

Students will create and answer quiz questions on the online quiz system Scuiiz (scuiz.org). You will need to go to this site and create an account for this class. You must create questions related to the course material in order to get access to questions to answer. Each correctly-answered question is worth one point. To get full credit for this course element, you will need to accumulate 180 points – about 20 answers per week, starting week 2. An additional 20 points will be accepted as extra credit. You get access to 10 questions when you create one, so you will need to create, on average, 2 questions per week. Note that the system automatically filters out questions that are too easy, so if you wait too long to participate, you will be left with only difficult questions and may have a hard time getting the needed points. You can also gain points by challenging questions that are incorrect or confusing.

Class sessions

We will alternate between lecture and lab settings, with lectures on Tuesdays and labs on Thursdays. Some lab sessions are designated data gathering or data analysis “specialist” labs. On these labs, attendance by the other technical specialists is optional. For example, data analysis specialists need not attend data gathering specialist labs. Project coordinators are required to attend all specialist labs, and should ensure that the appropriate other members of the team are present.

Some days are designated “project work days”. All group members are expected to attend on these days to ensure maximum joint work time.

Grading

Grading will be 40% individual and 60% group-based:

Individual:

- Individual lab exercises (3): 15%
- Journal entries (10): 10%
- Quizzes (180 total points): 10%
- Attendance / Participation: 5%

Group:

- Group assignments (5): 15%
- Specialist labs (4): 10%
- Group project (6 milestones + final report): 35%

Group contributions will be regularly evaluated by the instructor and your peers. Please pull your own weight!

Late Assignments

Lab assignments are due the day after the lab is held. Other assignments are due on their respective due dates given below. Each student has three late days without penalty. These can be used on any individual assignment, except the quizzes. Once the three days are used up, no more late assignments will be accepted. The group also has three late days for assignments and milestones, except for Milestone 6 (Status Report) and the final report.

Schedule

(Lab classes - **in bold** - meet in CDM 819)

9/6: Lab 0: Introduction

Course structure and expectations. Groups and responsibilities. Course tools: Slack, Jupyter notebook. Review of Python data structures.

9/11: Social networks

Social networks and social network analysis.

Reading: Golbeck, Ch. 1

9/13: Data and visualization

Pandas data frames. Simple plotting in matplotlib.

Group assignments announced.

In class: Lab 1

9/14 (and all subsequent Fridays through 11/9)

Journal due

9/18: Graph basics 1

Nodes and edges. Directed and undirected networks. Weighted and unweighted networks. Representing networks: edge list and adjacency matrix. Connectedness and connected components. Degree. Degree distribution. Graph file formats: Pajek, GML, GraphML.

Reading: Golbeck, Ch. 2

9/20: NetworkX

Using NetworkX to create and visualize small networks. Calculating node degree and degree distributions. Saving and loading networks.

In class: Lab 2

9/25: Graph basics 2

Node and edge attributes. Bipartite networks. Paths, geodesics. Diameter. Density. Characteristics of social networks.

Due: Group expectations agreement

Reading: Golbeck, Ch. 4

9/27: Data gathering 1

Using web APIs from Python. OAuth. Being a good citizen. Snowball sampling. Node, edge and attribute gathering. Storing network data.

In class: Specialist lab 1 (OAuth and APIs)

Due: Milestone 1 (Slack posts)

10/2: Centrality 1

Measuring node importance. Degree centrality. In- and out-degree. Eigenvector centrality and PageRank.

Due: Group assignment 1 (Degree distribution)

Reading: Golbeck, Ch. 3

10/4: Gephi

Using Gephi to visualize networks. Sizing, coloring and labeling nodes and edges. Layout algorithms.

In class: Lab 3

Due: Quiz 1

10/9: Centrality 2

Structural measures of importance: betweenness, closeness. Brokerage and structural holes.

Due: Group assignment 2 (Network visualization)

In class: Team progress report

Reading: Golbeck, Ch. 5

10/11: Data analysis 1

The matplotlib package in Python. Barplots and scatter plots. Performing analyses and visualizations on data extracted from networks.

Due: Milestone 2 (Initial data gathering)

In class: Specialist lab 2 (Matplotlib)

10/16: Bipartite networks

Special characteristics of bipartite networks. Computing and visualizing one-mode projections. Handling weighted projections. Edge filtering.

Due: Milestone 3 (Analysis plan)

Reading: Golbeck, Ch. 8

10/18: Group work time

In class: Informal peer evaluation

10/23: Community detection

Purposes of community detection. Modularity, betweenness, random walk, and other clustering methods. Overlapping communities and large-scale problems.

Due: Group assignment 3 (Projection)

10/25: Data gathering 2

Anonymization. Handling errors. Managing the crawling queue. Monitoring and managing long-duration processes.

In class: Specialist lab 3 (Snowball sampling)

Due: Quiz 2

10/30: Assortativity

Homophily in social networks. Degree and attribute assortativity. Assortativity metrics.

Reading: Golbeck, Ch. 7

11/1: Data analysis 2

Labeling and annotating plots in matplotlib. Advanced visualization in Gephi.

In class: Specialist lab 4 (Advanced visualization)

Due: Milestone 4 (Data)

11/6: Random graph models

Poisson random graphs, growing random graphs, preferential attachment processes and graph characteristics.

Due: Group assignment 4 (Bipartite projections)

11/8: Project work time

Due: Milestone 5 (Draft visualizations)

11/13: Project status report

Due: Milestone 6 (Status report)

In class: Presentations

11/15: No class

Due: Group assignment 4 (Betweenness)

11/16

Due: Quiz answers

11/20: Final report due

Formal peer evaluation due (counts as final journal entry)

Course Policies

Attendance

Students are expected to attend each class and to remain for the duration. Coming 15 minutes late or leaving 15 minutes early constitutes an absence for the student. Students are individually responsible for material they may have missed due to absence or tardiness.

Assignment Submission

All assignments will be submitted to D2L or in some cases, handed in during class. Do not submit assignments by email.

Late assignments

Each student has three late days that can be used without penalty for any individual assignment, except quizzes. After the three late days, no further late assignments will be accepted. Each team also has three late days that can be used for group assignments and project milestones, except Milestone 6 and the Final

Report. Individual late days cannot be used for team assignments, or vice versa. Only with documentation provided through the Dean of Students Absence Notification process can late assignments be considered outside of these limits.

Attitude

A professional and academic attitude is expected throughout this course. Measurable examples of non-academic or unprofessional attitude include but are not limited to: talking to others when the instructor is speaking, mocking another's opinion, cell phones ringing, emailing, texting or using the Internet whether on a phone or computer. If any issues arise a student may be asked to leave the classroom. The professor will work with the Dean of Students Office to navigate such student issues.

Civil Discourse

DePaul University is a community that thrives on open discourse that challenges students, both intellectually and personally, to be socially responsible leaders. It is the expectation that all dialogue in this course is civil and respectful of the dignity of each student. Any instances of disrespect or hostility can jeopardize a student's ability to be successful in the course. The professor will partner with the Dean of Students Office to assist in managing such issues.

Cell Phones/On Call

If you bring a cell phone to class, it must be off or set to a silent mode. Should you need to answer a call during class, students must leave the room in an undistruptive manner. Out of respect to fellow students and the professor, texting is never allowable in class. If you are required to be on call as part of your job, please advise me at the start of the course.

University Policies

Changes to Syllabus

This syllabus is subject to change as necessary during the quarter. If a change occurs, it will be thoroughly addressed during class, posted under Announcements in D2L and sent via email.

Online Course Evaluations

Instructor and course evaluations provide valuable feedback that can improve teaching and learning. The greater the level of participation, the more useful the results. As students, you are in the unique position to view the instructor over time. Your comments about what works and what doesn't can help faculty build on the elements of the course that are strong and improve those that are weak. Isolated comments from students and instructors' peers may also be helpful, but evaluation results based on high response rates may be statistically reliable (believable). As you experience this course and material, think about how your learning is impacted. Your honest opinions about your experience in and commitment to the course and your learning may help improve some components of the course for the next group of students. Positive comments also show the department chairs and college deans the commitment of instructors to the university and teaching evaluation results are one component used in annual performance reviews (including salary raises and promotion/tenure). The evaluation of the instructor and course provides you an opportunity to make your voice heard on an important issue – the quality of teaching at DePaul. Don't miss this opportunity to provide feedback!

Academic Integrity and Plagiarism

This course will be subject to the academic integrity policy passed by faculty. More information can be found at <http://academicintegrity.depaul.edu/>.

The university and school policy on plagiarism can be summarized as follows: Students in this course should be aware of the strong sanctions that can be imposed against someone guilty of plagiarism. If proven, a charge of plagiarism could result in an automatic F in the course and possible expulsion. The strongest of sanctions will be imposed on anyone who submits as his/her own work any assignment which has been prepared by someone else. If you have any questions or doubts about what plagiarism entails be sure to consult the instructor. While students are permitted to discuss assignments at the conceptual level, under no circumstances should students share specific answers (electronically or otherwise).

Withdrawal

Students who withdraw from the course do so by using the Campus Connection system (<http://campusconnect.depaul.edu>). Withdrawals processed via this system are effective the day on which they are made. Simply ceasing to attend, or notifying the instructor, or nonpayment of tuition, does not constitute an official withdrawal from class and will result in academic as well as financial penalty.

Retroactive Withdrawal

This policy exists to assist students for whom extenuating circumstances prevented them from meeting the withdrawal deadline. During their college career students may be allowed one medical/personal administrative withdrawal and one college office administrative withdrawal, each for one or more courses in a single term. Repeated requests will not be considered. Submitting an appeal for retroactive withdrawal does not guarantee approval. College office appeals for CDM students must be submitted online via MyCDM. The deadlines for submitting appeals for this quarter is the last day of the last final exam of Winter Quarter 2019.

Excused Absence

In order to petition for an excused absence, students who miss class due to illness or significant personal circumstances should complete the Absence Notification process through the Dean of Students office. The form can be accessed at <http://studentaffairs.depaul.edu/dos/forms.html>. Students must submit supporting documentation alongside the form. The professor reserves the sole right whether to offer an excused absence and/or academic accommodations for an excused absence.

Incomplete

An incomplete grade is a special, temporary grade that may be assigned by an instructor when unforeseeable circumstances prevent a student from completing course requirements by the end of the term and when otherwise the student had a record of satisfactory progress in the course. CDM policy requires the student to initiate the request for incomplete grade before the end of the term in which the course is taken. Prior to submitting the incomplete request, the student must discuss the circumstances with the instructor. Students may initiate the incomplete request process in MyCDM.

- All incomplete requests must be approved by the instructor of the course and a CDM Associate Dean. Only exceptional cases will receive such approval.
- If approved, students are required to complete all remaining course requirement independently in consultation with the instructor by the deadline indicated on the incomplete request form.
- By default, an incomplete grade will automatically change to a grade of F after two quarters have elapsed (excluding summer) unless another grade is recorded by the instructor.
- An incomplete grade does NOT grant the student permission to attend the same course in a future quarter.

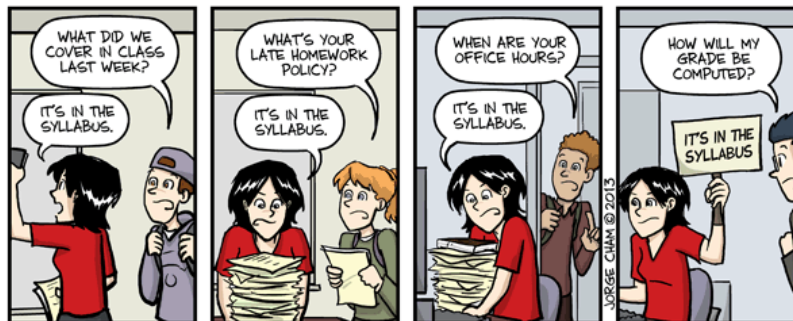
Students with Disabilities

Students who feel they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss their specific needs. All discussions will remain confidential. To ensure that you receive the most appropriate accommodation based on your needs, contact the instructor as early as possible in the quarter (preferably within the first week of class), and make sure that you have contacted the Center for Students with Disabilities (CSD) at: Student Center, LPC, Suite #370 Phone number: (773)325.1677 Fax: (773)325.3720 TTY: (773)325.7296

Quarter at a Glance

Lecture	Lab	Specialist Lab

Date	Topic	Major Assignments	Project
9/6	Lab introduction		
9/11	Introduction		
9/13	iPython lab		Groups announced
9/18	Graph basics 1		
9/20	NetworkX lab		
9/25	Graph basics 2		Group policies
9/27	Data gathering lab 1		
10/2	Centrality 1	Group assignment 1	
10/4	Gephi lab		Milestone 1
10/9	Centrality 2	Group assignment 2	
10/11	Data analysis lab 1		Milestone 2
10/16	Bipartite networks		
10/18	Group work time		Milestone 3
10/23	Community detection		
10/25	Data gathering lab 2		
10/30	Assortativity	Group assignment 3	
11/1	Data analysis lab 2		Milestone 4
11/6	Random graphs		
11/8	Project work time		Milestone 5
11/13	Status report		
11/15	No class	Group assignment 4	Milestone 6
11/16	No class		
11/20	Final report		Final report



IT'S IN THE SYLLABUS

This message brought to you by every instructor that ever lived.

WWW.PHDCOMICS.COM