CSC 299: Sophomore Lab in Applied Computing
"Mining Social Networks"

Fall 2014
Tuesdays: 11:50 – 1:20 pm CDM 819
Thursdays: 11:50 – 1:20 pm, CDM 216
Professor Robin Burke
Office hours: Tuesdays 10:00 – 11:30 (CDM 849) and 2:30 – 4:00 pm (Lincoln Park, 990 W. Fullerton, Suite 3100) and by appointment.
Contact: (Phone) 312-362-5910
(Email - better) rburke@cs.depaul.edu
(Piazza – best) www.piazza.com

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

iPython Notebook

We will use iPython 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 iPython 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. Be aware: the latest public release of Gephi is not compatible with Java 8. Until the developers complete the next release of Gephi, you will need to have a Java 7 SDK on your machine to run Gephi. Various work-around may be available for your particular platform. Feel free to complain to the Gephi development team about this.

Piazza

We will be using Piazza for class discussion. The system is designed to get you help quickly and efficiently from your classmates and the instructor. Rather than emailing questions, I encourage you to post your questions on Piazza. Find our class page at: https://piazza.com/depaul/fall2014/csc299/home

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 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.
- 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 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 student reflect on their group project work and to document steady progress towards project goals.

Class sessions

We will alternate between lecture and lab settings, with labs on Tuesdays and lectures 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.

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

Note: I will be out of town during the week of October 7. The lab that would ordinarily be on October 7 will be held instead on the preceding Thursday, October 2. The two lectures on centrality measures will be presented by a guest instructor.

Grading

Grading will be 25% individual and 75% group-based:

Individual:

- Individual lab exercises (3): 15%
- Journal entries (10): 10%

Group:

- Group assignments (5): 15%
- Specialist labs (4): 10%
- Group project (6 milestones + final report): 50%

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. Late assignments will be marked off by 10% per day, up to 3 days late.

Schedule

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

9/11: Introduction

Course structure and expectations. Course tools. Groups and responsibilities. Ethical uses of social network information.

Reading: Golbeck, Ch. 1

9/16: iPython

The iPython environment. Loading and saving notebooks. Installing and importing Python libraries. List comprehensions.

In class: Lab 1

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.

Group assignments announced.

Reading: Golbeck, Ch. 2

9/23: NetworkX

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

In class: Lab 2

Due: Group expectations agreement

9/25: Graph basics 2

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

Due: Group assignment 1 (Degree distribution)

Reading: Golbeck, Ch. 4

9/30: Gephi
Using Gephi to visualize networks. Sizing, coloring and labeling nodes and edges. Layout algorithms.
In class: Lab 3

10/2: Data gathering 1 (Note special Thursday lab time)
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: Group assignment 2 (Network visualization)

10/7: Centrality 1 (Note no lab this week)
Measuring node importance. Degree centrality. In- and out-degree. Eigenvector centrality and PageRank.
Due: Milestone 1 (Support web page)
Reading: Golbeck, Ch. 3

10/9: Centrality 2
Structural measures of importance: betweenness, closeness. Brokerage and structural holes.
Reading: Golbeck, Ch. 5

10/14: 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: Group assignment 3 (Betweenness)
Reading: Golbeck, Ch. 8

10/21: Project work time
Due: Milestone 3 (Analysis plan)
In class: Informal peer evaluation

10/23: Random graph models
Poisson random graphs, growing random graphs, referential attachment processes and graph characteristics.
Due: Group assignment 4 (Bipartite projections)

10/28: Data gathering 2
Anonymization. Handling errors. Managing the crawling queue. Monitoring and managing long-duration processes.
In class: Specialist lab 3 (Snowball sampling)

10/30: Community detection
Purposes of community detection. Modularity, betweenness and other clustering methods. Overlapping communities and other open problems.
Due: Milestone 4 (Data)

11/4: Data analysis 2
Labeling and annotating plots in matplotlib. Advanced visualization in Gephi.
In class: Specialist lab 4 (Advanced visualization)

11/6: Assortativity
Homophily in social networks. Degree and attribute assortativity. Assortativity metrics.
Due: Milestone 5 (Draft visualizations)

Reading: Golbeck, Ch. 7

11/11: Progress presentations / Project work time
Due: Milestone 6 (Status report)

11/13: Recommender systems
User-item associations. Node similarity via correlation. Recommender systems.
Reading: Golbeck, Ch. 13

11/18: Project work time

11/25: 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.

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 undisruptive 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 2014.

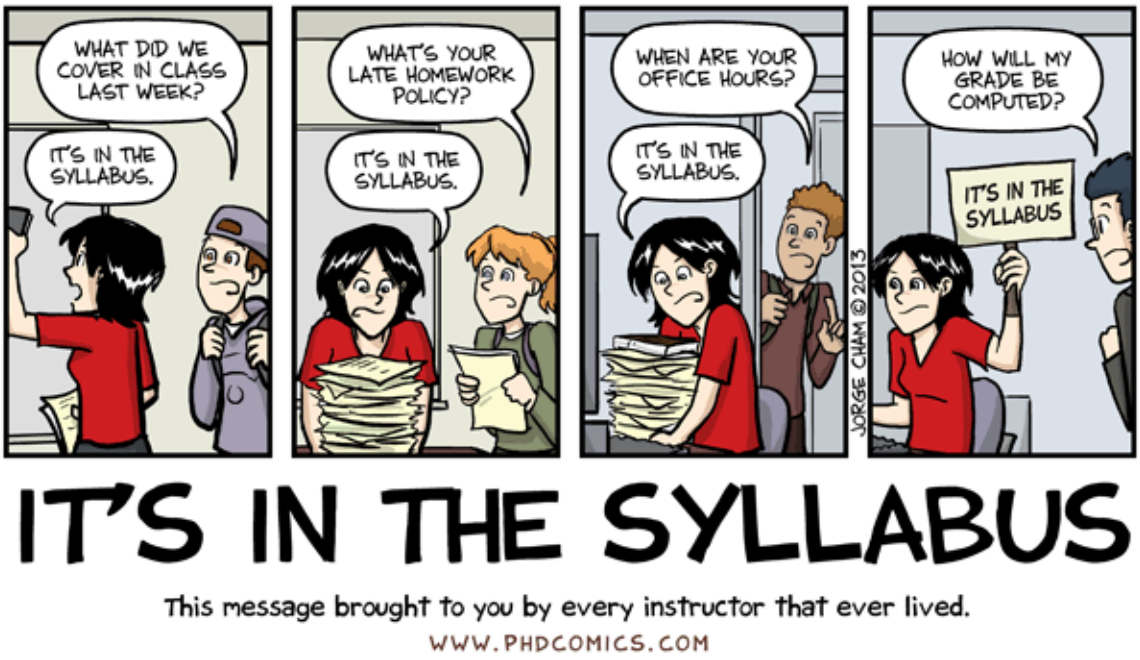
Excused Absence
In order to petition for an excused absence, students who miss class due to illness or significant personal circumstances should complete the Absence Notification process through the Dean of Students office. The form can be accessed at http://studentaffairs.depaul.edu/dos/forms.html. 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	Assignments	Project
9/22	Introduction		
9/16	iPython lab		
9/18	Graph basics 1		Groups announced
9/23	NetworkX lab		Group expectations
9/25	Graph basics 2	Group assignment 1	
9/30	Gephi lab		
NOTE: THURSDAY LAB DAY			
10/2	Data gathering lab 1	Group assignment 2	
NOTE: TUESDAY LECTURE DAY			
10/7	Centrality 1		Milestone 1
10/9	Centrality 2		
10/14	Data analysis lab 1		Milestone 2
10/16	Bipartite networks	Group assignment 3	
10/21	Project work time		Milestone 3
10/23	Random graphs	Group assignment 4	
10/28	Data gathering lab 2		
10/30	Community detection		Milestone 4
11/4	Data analysis lab 2		
11/6	Assortativity		Milestone 5
11/11	Status reports Project work time		Milestone 6
11/13	Recommender systems		
11/18	Project work time		
11/25	Final report		Final report