CSC 495: Social Network Analysis

Professor Robin Burke
Fall 2018, Section 701/710, TBA
Thursdays 5:45 — 9:00 pm

Office: (Loop) CDM 841, (Lincoln Park) 990 W. Fullerton, Rm 3130.
Course discussion on Slack (recommended): https://csc495fall18.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

This course is an introduction to the concepts and methods of social network analysis. Students will learn to extract and manage data about network structure and dynamics, and to analyze, model and visualize such data. Students will use software tools to model and visualize network structure and dynamics. Specific network applications to be discussed include online social networks, collaboration networks, and communication networks. PREREQUISITE(S): CSC 423 or CSC 400 or SOC 412

Kolaczyk, E. and Csardi, G. *Statistical Analysis of Network Data with R*. Springer. ISBN: 978-14939-0982-7. Other readings online as assigned including:

"Social Network Data". Chapter 1 in Hanneman, Robert A. and Mark Riddle. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/)
Hunter, David R., et al. "ergm: A package to fit, simulate and diagnose exponential-family models for networks." *Journal of statistical software* 24.3 (2008): nihpa54860.
Liben-Nowell, David, and Jon Kleinberg. "The link-prediction problem for social networks." *Journal of the Association for Information Science and Technology* 58.7 (2007): 1019-1031.
Malewicz, Grzegorz, et al. "Pregel: a system for large-scale graph processing." *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, 2010.
Robins, Garry, et al. "An introduction to exponential random graph (p*) models for social networks." *Social networks* 29.2 (2007): 173-191.
Xin, Reynold S., et al. "GraphX: Unifying data-parallel and graph-parallel analytics." *arXiv preprint* arXiv:1402.2394 (2014).

Learning Objectives

After taking this class, students will be able to:

- Define key concepts in networks / graphs such as node, edge, path, geodesic, weighted and directed networks.
- Apply multiple representations of networks including edge list, adjacency list and adjacency matrix.
- Use linear algebra techniques such as matrix manipulation to compute network operations and properties.
- Identify appropriate relations to encode as edges and attributes in a data set.
- Define and calculate key metrics such as degree, degree distribution, diameter, density, closeness, betweenness, modularity, local and global clustering.
- Interpret the results of computing various forms of centrality including degree, closeness,

betweenness, eigenvector, and PageRank centrality.

- Explain and distinguish between several key network models: random, growing random, preferential attachment, and exponential random graph.
- Compute and analyze projections of bipartite networks and employ filtering techniques to such networks.
- Analyze different types of networks to identify important actors, significant structural features, such as communities, and overall network characteristics.
- Describe the consequences of different network structures for network processes including contagion and social influence.
- Use R to load, process, filter, manipulate and compute metrics over a variety of networks.
- Apply QAP regression and exponential random graph modeling to test hypotheses about
- graph formation.
- Identify key differences between distributed models of graph computation including edge-parallel and vertex-parallel.
- Identify different supervised and unsupervised link prediction models.
- Create effective network visualizations that communicate analytic findings about a network.

Assessment

Students will be assessed on the basis of 6 homework assignments, 2 labs, online quiz questions, and a final project.

Participation: 5%
Homework (6): 30% (extra credit up to 2.5%)
Labs (2): 10%
Quiz questions: 25% (extra credit up to 3%)
Final project: 30%

Homeworks are due weekly at 11:59 pm on the day (Friday) following class nights. Attendance at labs is optional, and lab assignments will be due the following week. Lab sessions will be recorded so on-line students can follow along to complete each exercise.

Some classes will involve an extended worked example in which we work through a network analysis problem in class. Students can submit their own coding of these examples for extra credit, totaling up to 2.5% bonus towards the homework grade.

Students will create and answer quiz questions on the online quiz system Scuiz (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.

The final project will involve the creation of a network analysis using a data set of the student's choice. There are seven milestones:

M1: Project brainstorming (due 4/10)
M2: Project proposal (4/24)
M3: Project data (5/1)
M4: First visualization (5/8)
M5: Visualization peer review (5/15)
M6: Draft visualizations (5/22)
M7: Project due (6/5)

Tentative Schedule

9/6: Introduction

Introduction to the class. Syllabus and expectations. Basic terminology and concepts. The igraph package in R.
Reading: SAND, Ch. 1 and 2
Online: R review, ggplot intro, Christakis TED talk

9/13: Bipartite Networks / Network Data

Bipartite networks and projections. Defining networks. Gathering network data. Network representations.
Reading: Hannemen and Riddle, *Introduction to social network methods*, Ch. 1, "Social network data".
Due: Homework 1, Extra credit 1

9/20: Centrality Measures

Centrality measures: degree, betweenness, closeness. Eigenvector-based centrality measures: Katz, PageRank, leading eigenvector.
Reading: SAND, Ch. 4.1, 4.2
Due: Homework 2, EC 2, Project brainstorming

9/27: Network Visualization / Lab

Principles of visualization as applied to networks. Size, color and layout. Introduction to the course project.
Lab: Loading, manipulating, and visualizing network data in Gephi.
Reading: SAND, Ch. 3
Due: Homework 3, EC 3,

10/4: Network Structure

Local structure. Community detection algorithms.
Reading: SAND, Ch. 4.4 – 4.6
Due: Homework 4, Lab 1, Project proposal

10/11: Mathematical Models

Poisson random graphs. Growing random networks. Preferential attachment. Properties and phase transitions. Degree distributions. QAP regression.
Reading: SAND, Ch. 5
Due: Homework 5, EC 4, Project data

10/18: Statistical Models I

Frameworks for evaluating results in network analysis: autocorrelation, matching techniques, QAP regression, and other models. Computational considerations. Exponential random graph modeling. Interpreting ERGM results.
Reading: SAND, Ch. 6.1 – 6.3; Robins, et al. 2007 "An introduction to exponential random graph (p*) models for social networks".
Due: First visualization, EC 5

10/25: Statistical Models II

Formulating ERGM models. Controlling the fitting process. Goodness of fit calculations. Strategies for building good models. Lab: Applying ERGM analysis.
Reading: Hunter, et al. 2008 "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks".
Due: Visualization peer review, EC 6

11/1: Large-Scale Graph Analysis

Scalability and parallelism in graph analytics. MapReduce as applied to graph calculations and current extensions including Pregel and GraphX.
Reading: Malewicz, et al. 2010 "Pregel: A System for Large-Scale Graph Processing"; Xin, et al. 2014 "GraphX: Unifying Data-Parallel and Graph-Parallel Analytics"
Due: Draft visualizations, Lab 2

11/8: Link Prediction / Lab

Supervised and unsupervised link prediction techniques. Applications. Network-oriented recommender systems. Lab: Project work
Reading: SAND, Ch. 7.1 and 7.2; Liben-Nowell & Kleinberg, 2007 "The Link-Prediction Problem for Social Networks"
Due: Homework 6

11/15: Final project due

11/16: Quiz questions

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. Do not submit assignments by email.

Late assignments

Each student has three late days for homeworks and labs. These can be used without penalty on any assignment. After the three days have been used, assignments will not be accepted late. Each student also has three late days for the final project, which can be used for any milestone, except those that require peer response (Milestones 1, 4 and 5) and the final report. Late days cannot be moved between the two categories. 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 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. 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. You will receive an email notification when the course evaluations are ready to be completed.

Academic Integrity and Plagiarism

This course will be subject to the University's academic integrity policy. 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 Fall Quarter 2017.

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

Week	Date	Topic	Reading	Due
1	9/6	Introduction	SAND, Ch. 1 and 2	
2	9/13	Bipartite Networks Network Data	Hanneman and Riddle, Ch. 1	H1, EC1
3	9/20	Centrality	SAND, Ch. 4.1 and 4.2	H2, EC2, M1
4	9/27	Network Visualization Lab	SAND, Ch. 3	H3, EC3
5	10/4	Community Detection	SAND, Ch. 4.4 – 4.6	H4, L1, M2
6	10/11	Mathematical Models	SAND, Ch. 5	H5, EC4, M3
7	10/18	ERGM 1	SAND, Ch. 6.1 – 6.3; Robins, et al. 2007	EC5, M4
8	10/25	ERGM 2 Lab	Hunter, et al. 2008	EC6, M5
9	11/1	Large-scale	Malewicz, et al. 2010; Xin, et al. 2014	L2, M6
10	11/8	Link Prediction Lab	SAND, Ch. 7.1 and 7.2; Liben-Nowell & Kleinberg, 2007	H6
Finals	11/15 11/16	No class		M7 Quiz