Course Name and Number: IS 620 Web Analytics
Credits: *3 cr.*
Prerequisite(s): IS 606 and IS 607

Course Summary:

Organizations, both commercial and community, can benefit from deep analysis of their website interactions and mobile data. Social networks have also become a source of information for companies; search engines are an important referral mechanism. Popular social networks and other online communities provide rich sources of user information and (inter-) actions through their application programming interfaces. This data can help to identify a number of individual user preferences and behaviors, as well as fundamental relationships within the community. Search engines use algorithms to rank sites. Students will learn how to analyze social network data for types of networks, the fundamental calculations used in social networks (e.g., centrality, cohesion, affiliations, and clustering coefficient) as well as network structures and roles. Beyond social network data, students will learn about important concepts of analyzing website traffic such as click streams, referrals, keywords, page views, and drop rates. The course will touch on the fundamentals of search algorithms and search engine optimization. To provide a basic context for understanding these online user and community behaviors, students will learn about relevant social science theories such as homophily, social capital, trust, and motivations as well as business and social use contexts. In addition, this course will address ethical and privacy issues as they relate to information on the Internet and social responsibility.

Course Learning Outcomes:

At the end of this course, students will be able to:

- Analyze text data, including natural language processing and text representation, word association, topic mining, opinion mining and sentiment analysis, and text-based prediction.
- Perform network analysis, including creating graphs, calculating statistics on nodes, and graph visualization.
- Work with various social network APIs, including Twitter, Facebook, and Linked In.

Students will be required to:

- Apply what they learning about network analysis and text mining in a series of increasing complex projects and associated presentations.

How is this course relevant for data analytics professionals?

Text mining is about working with unstructured data. Network analysis focuses more on relationships than entities. These are two of the fastest growing sub-fields of data science, and are increasingly successful for success in the workplace.

Assignments and Grading:

Assignments (8 x 25)	20%
Projects (4 x 100)	40%
Final Project (1 x 200)	20%
Final Project Presentation (1 x 50)	5%
Discussion Participation (15 x 10)	15%
TOTAL	100%

Grades:

Quality of Performance	Letter Grade	Range %	GPA/ Quality Pts.
Excellent - work is of exceptional quality	A	93 - 100	4.0
	A-	90 - 92.9	3.7
Good - work is above average	B+	87 - 89.9	3.3
Satisfactory	B	83 - 86.9	3.0
Below Average	B-	80 - 82.9	2.7
Poor	C+	77 - 79.9	2.3
	C	70 - 76.9	2.0
Failure	F	< 70	0.0

Required Texts and Materials:

- *Mining the Social Web, 2/e*, Matthew A. Russell, Oct 20, 2013. https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition provides a rich set of additional materials; author's blog http://miningthesocialweb.com/
- *Social Network Analysis for Startups*, Maksim Tsvetovat and Alexander Kouznetsov, O'Reilly, Sep 30, 2011.
- *Natural Language Processing with Python*, Steven Bird, Ewan Klein, and Edward Loper, O'Reilly, Jun 30, 2009.

Other reading material (all freely available on-line)

- *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, David Easley and Jon Kleinberg, Cambridge 2010. Freely downloadable: http://www.cs.cornell.edu/home/kleinber/networks-book/
- *Network Science Book,* Lazlo Barabasi. Freely downloadable: http://barabasilab.neu.edu/networksciencebook/. Printed version expected 2015 by Cambridge University Press
- *Graph Databases*, Ian Robinson, Jim Webber, and Emil Eifrem, O'Reilly, June 20, 2013. Freely downloadable from O'Reilly.

- *Mining of Massive Datasets*, Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman, http://infolab.stanford.edu/~ullman/mmds/book.pdf. Supporting materials, including links to Coursera course: http://www.mmds.org/

Relevant Software, Hardware, or Other Tools:

- Python 2.7 (free distribution from Anaconda here: https://www.continuum.io/downloads)
- GraphLab Create. Freely available for students
- Assignments turned in as IPython notebooks, with notebooks in github, and links to notebooks in assignment submission text.

My Contact Information:

Alain Ledon

cuban.quant@gmail.com

You are encouraged to ask me questions on the "Ask Your Instructor" forum on the course discussion board where other students will be able to benefit from your inquiries.
I am available by e-mail or by cell phone. We can set up virtual one-on-one meetings. For the most part, you can expect me to respond to questions by email within 24 to 48 hours. If you do not hear back from me within 48 hours of sending an email, please resend your message."

Course Outline:

Unit	Topics	Readings	Deliverables
Week #1 **Aug 29-Sep 5**	**Set up Environment**	Supplementary materials on Gephi and GraphLab Create	**Environment Setup**
Week #2 **Sep 6-Sep 12**	**Network Analysis: Overview** **Text Mining: Overview**	*Natural Language Processing with Python*, Chapters 1 and 2. *Social Network Analysis for Startups*, Chapter 1 Supplementary materials on iGraph package.	**Week 2 Assignment**
Week #3 **Sep 13-Sep 19**	**Network Analysis:** **Graph Theory, Definitions**	*Social Network Analysis for Startups*, Chapter 2 Supplementary material on Graph Theory.	**Week 3 Assignment**
Week #4 **Sep 20-Sep 26**	**Network Analysis:** **Centrality Measures**	*Social Network Analysis for Startups*, Chapter 3	**Project 1**
Week #5 **Sep 27-Oct 3**	**Network Analysis:** **Clustering 1**	*Social Network Analysis for Startups*, Chapter 4	**Week 5 Assignment**
Week #6	**Network Analysis:**	*Social Network Analysis for Startups*,	**Project 2**

Unit	Topics	Readings	Deliverables
Oct 4-Oct 10	2-mode networks	Chapters 5 and 6	
Week #7 Oct 11-Oct 17	Text Mining: Natural Language Processing	*Natural Language Processing with Python*, Chapter 3 and 4.	Week 7 Assignment
Week #8 Oct 18-Oct 24	Text Mining: Word Association	*Natural Language Processing with Python*, Chapters 5 and 6.	Week 8 Assignment
Week #9 Oct 25-Oct 31	Network Analysis: Topic Mining 1	*Natural Language Processing with Python*, Chapters 7-8.	Project 3
Week #10 Nov 1-Nov 7	Network Analysis: Topic Mining 2	*Natural Language Processing with Python*, Chapter 9.	Week 10 Assignment
Week #11 Nov 8-Nov 14	Network Analysis: Sentiment Analysis	*Natural Language Processing with Python*, chapters 10 and 11.	Week 11 Assignment
Week #12 Nov 15-Nov 21	Text Mining: Text-Based Prediction	*Natural Language Processing with Python*, chapter 6. Supplementary material on algorithms	Week 12 Assignment
Week #13 Nov 22-Nov 28 *(Thanksgiving)*			
Week #14 Nov 29-Dec 5	Network Analysis and Text Mining: Longitudinal Analysis	TBD	Project 4
Week #15 Dec 6-Dec 12	Network Analysis and Text Mining: Working in the Cloud	TBD	Final Project Proposals Due
Week #16 Dec 13-Dec 21 (last day of classes)	Network Analysis and Text Mining: Working with Big Data	TBD	Final Projects and Presentations Due

How This Course Works

Meet-ups on Thursdays at 8:15 p.m.; Assignments due on following Sundays!

This course is conducted entirely online. Here is what your weekly workload and deliverable schedule will look like:

- Each week's material is available after 5:00 p.m. EDT on the Friday before that week's meet-up.
- You'll have a list of readings. There will also be a number of short videos to watch most weeks.

- There is a short, lightly graded discussion topic each week. Your initial post due before the meet-up, and your response due end of day the following Friday.
- Course meet-ups each Thursday at 8:15 p.m. Call in details below.
- For each course track, you'll submit four **projects** (due end of day EDT on the Sunday after the meet-up) and a **final project** (with a short **presentation** of your final project). Most weeks when there are not projects due, you'll have shorter coding assignments.
- You may always propose in advance to substitute your own datasets for the assigned datasets.
- Students are expected to complete all assignments by their due dates. Any work turned in after the due date will receive a maximum score of 80%. If solutions have been posted for an assignment before you've turned it in, you'll need to propose an alternative assignment acceptable to the instructor. Future data scientists please take note: there is an overwhelmingly positive correlation between how early students turn in their assignments and their course grades!
- There will also be short ungraded "hands on labs" that will help you prepare for your assignments.
- Working in teams on the projects is strongly encouraged, but not required. The ability to work effectively on virtual teams is an important "soft skill" for data scientists.
- If you take non-trivial amounts of code from the web or other sources, you must provide full attribution. This way, your grade will be based on the code that you added to the found "starter" code.

Meet-up Call-in Details

ACCESSIBILITY AND ACCOMMODATIONS

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: http://sps.cuny.edu/student_services/disabilityservices.html

ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please
see: http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

ACADEMIC INTEGRITY

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please
see: http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

STUDENT SUPPORT SERVICES

If you need any additional help, please visit Student Support Services:
http://sps.cuny.edu/student_resources/