CIS 600 Principles of Social Media and Data Mining
Fall 2020
Syllabus

Instructor: Edmund Yu, PhD
Associate Teaching Professor
Email: esyu@syr.edu
Phone: (315)443-3110
Office: CST 4-206L-B

Class meetings: Tuesday/Thursday 9:30am – 10:50am, Watson Theater
Note: Due to social distancing, this classroom can accommodate only about 36 students. Therefore, tentatively, for Tuesday classes, this classroom is reserved for students whose last names begin with A through M. For Thursday classes, it's reserved for students whose last names begin with N through Z. For those students who cannot or should not come to the classroom, please log on to the Blackboard and use Blackboard Collaborate Ultra to attend the class. I will use Blackboard Collaborate Ultra to record the lectures as well.

Office hours: Monday/Wednesday 10:00am – 12:00noon

TA Office hours: Pavan Kumar Vaddineni (pvaddine@syr.edu): Monday 12-2pm

Texts: The following books are recommended but not required:

- Mining the Social Web, Matthew Russell, 2019, O'Reilly, 3rd Edition, ISBN: 1491985046.
- Social Media Mining: An Introduction, by Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, 2014, Cambridge University Press, ISBN: 1107018854. (Pre-publication draft is available at: http://dmml.asu.edu/smm/book/)
- Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, by Bing Liu, Cambridge University Press, 2015, ISBN: 1107017890.

In addition, parts of the following books will be referenced:

- Community Detection and Mining in Social Media, Tang & Liu, 2010, Morgan and Claypool Publishers. (Can be downloaded from its publisher, Morgan & Claypool, http://www.morganclaypool.com/doi/abs/10.2200/S00298ED1V01Y201009DMK003, if on campus.)
- Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Jon Kleinberg, 2010, Cambridge University Press (online: http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf)
- Natural Language Processing with Python, 2nd Edition, Steven Bird et al. (not yet released, but online preview version is available at http://www.nltk.org/book/.

Additional materials will also be made available via Blackboard.

Course description:
The past two decades have witnessed the rapid rise of social media. Founded in February 2005 and bought by Google in October 2006 for $1.65 billion dollars, the video-sharing site YouTube, in its 20 months of existence, had grown from zero to more than 20 million unique user visits per day, dishing out 100 million video views every day, making it a textbook case of 'viral marketing' - spreading from person to person similar to the way the pandemic flu spreads. Founded in February 2004, the mega social

networking site Facebook surpassed Google in March 2010 to become the most visited website in the US. The sheer size of its user base, over 3.14 billion active users currently, has formed a new 'World of Mouth,' to borrow a term from the wildly popular book Socialnomics by Erik Qualman, for 'word-of-mouth marketing.' This still ongoing development has already greatly affected how people make friends, get information, do business, and even start a revolution.

Social media mining is a branch of data mining, with its focus set on social media. Simply put, data mining refers to the process of discovering (interesting/actionable) patterns, or extracting knowledge, from data. Accordingly, social media mining refers to the process of discovering patterns, or extracting knowledge, from social media. In this course, we will introduce you to popular methods for discovering pattern and extracting knowledge from social media. When introducing those methods, we will balance theory and practice. We will explain the key concepts and theoretical foundations behind those methods, but the students will also acquire hands-on experiences using publicly available software to develop applications for their own scientific or business problems or just for fun.

Prerequisites: Proficiency in object-oriented programming

Course Objectives:
Students should fully understand the technical issues, including key concepts, theoretical foundations and basic algorithms, involved in discovering patterns and extracting knowledge from social media. Students will also acquire tremendous amount of hands-on experiences in developing social media mining applications for their own scientific or business problems, or simply for fun.

Course Outcomes:
After completion of the course, students should be able to:

1. Understand the state of the art of the six types of social media
2. Harvest social media data using social network APIs, such as Twitter APIs, and Facebook Graph API, together with the Python programming language
3. Understand the key concepts and principles behind social networks analysis tasks such as centrality analysis, influence modeling and maximization, and community detection
4. Programmatically analyze social networks, including their structures and contents, as well as user attributes and behaviors
5. Programmatically analyze the sentiments (or emotional contents) contained in tweets, using either lexicon-based approach or machine learning algorithms
6. Develop social media mining applications.

Class Schedule:

Week 1: Introduction: Social Media and Social Media Mining (August 25)
Python Tutorial (August 27)
Week 2: Python Tutorial, continued (September 1)
Harvesting Social Media Data Using Twitter Search API (September 3, 8)
Homework 1: Set up your Twitter app. Experiment with Twitter Search API.

Assignment 1: Programming in Python: due September 11 (10 points)

Week 3: Harvesting Social Media Data Using Twitter REST API (September 10)
Homework 2: Experiment with Twitter REST API.
Week 4: Harvesting Social Media Data Using Twitter Streaming API (September 15)

Homework 3: Experiment with Twitter Streaming API.
Social Network Analysis, Part 1: Introduction to Graph Theory & Networkx (September 17)

Assignment 2: Characterize your Twitter social network: due October 2 (10 points)

Week 5: Social Network Analysis, Part 2: Small Worlds & Six Degrees of Separation (September 22)
Social Network Analysis, Part 3: Network Models (September 24)
Week 6: Social Network Analysis, Part 4: Centrality Analysis (September 29)

Exam #1 (October 1, via Blackboard): 20 points

Week 7: Social Network Analysis, Part 5: Influence Modeling (October 6)
Social Network Analysis, Part 6: Community Detection (October 8)
Week 8: Data Mining Essentials (October 13, 15)
Week 9: Sentiment Analysis, Part 1: An Overview (October 20, 22)

Term Project Proposal: due October 25 (10 points)

Week 10: Sentiment Analysis, Part 2: Classification Algorithms & Tools (October 27, 29)
Week 11: Sentiment Analysis, Part 3: NLP & NLTK (November 3, 5)
Week 12: Harvesting Social Media Data Using Other APIs (November 10)
Social-media based Recommendation Systems (November 12)

Week 13: ***Term Project Presentations/Demonstrations (Nov 17, 19)***

Exam #2 (November 24, via Blackboard): 20 points

Term Project Deliverables: due December 11, 2020 (30 points)
A zipped folder that contains the following:

- The PowerPoint presentation
- A report that describes your approach, data, experimental results and analysis, plus anything else that you think worth mentioning. 3 pages per team member. (Single-spaced, with fonts no larger than 12 points.)
- Source code (with comments), data files, URL links, etc.
- A README file that describes how to install, run and use your system. If your code is available online, please mention the URL in the README file.

Grading:
Assignments 1 & 2: 20 points
Exams: 40 points (20 points each)
Term Project: 40 points (10 points each for the proposal, presentation, report, and code)

Note: A grade of incomplete will not be given, except under very extenuating circumstances and at the discretion of the instructor.

Attendance:
Attendance in class is very important. During each lecture period, information regarding assignments, due dates, explanation and clarification of assignments, and material that is not covered in the textbooks will be presented. If you miss a class for any reason, it is your responsibility to become familiar with the missed material. (My own slides will be uploaded to Blackboard.) Be prepared to spend extra time each week on this class, outside of the classroom.

ABET:
As part of the regular ABET accreditation process for the undergraduate program in computer science, we will be collecting samples of students' work in each of our undergraduate classes. As a result, some of your labs/homeworks/exams may be photocopied and saved to present to the ABET evaluators who visit next fall.

Academic Integrity:
"Syracuse University sets high standards for academic integrity. Those standards are supported and enforced by students, including those who serve as academic integrity hearing panel members and hearing officers. The presumptive sanction for a first offense is course failure, accompanied by the transcript notation "Violation of the Academic Integrity Policy." The standard sanction for a first offense by graduate students is suspension or expulsion. Students should review the Office of Academic Integrity online resource "Twenty Questions and Answers About the Syracuse University Academic Integrity Policy" and confer with instructors about course-specific citation methods, permitted collaboration (if any), and rules for examinations. The Policy also governs the veracity of signatures on attendance sheets and other verification of participation in class activities. Additional guidance for students can be found in the Office of Academic Integrity resource: 'What does academic integrity mean?'"

Faith Tradition Observances:
Syracuse University recognizes the diverse faith traditions represented among its campus community and supports the rights of faculty, staff, and students to observe according to these traditions.
Faculty are asked to make appropriate accommodation for students' observance needs by providing an opportunity to make up any examination, study, or work requirement that is missed because of an absence due to a religious observance, provided the instructor has been notified no later than the end of the second week of classes. No fees will be charged to the student for the costs incurred by the University for such make-up work. If a faculty member is unwilling or unable to make an appropriate accommodation, the student should consult his or her academic dean. SU's religious observances policy can be found at: http://supolicies.syr.edu/emp_ben/religious_observance.htm.

Disability Services:
If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), http://disabilityservices.syr.edu, located in Room 309 of 804 University Avenue, or call (315) 443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented Disabilities Accommodation Authorization Letters, as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible.

Syracuse University and I are committed to your success and to supporting Section 504 of the Rehabilitation Act of 1973. This means that in general no individual who is otherwise qualified shall be excluded from participation in, be denied benefits of, or be subjected to discrimination under any program or activity, solely by reason of having a disability. You are also welcome to contact me privately to discuss your academic needs although I cannot arrange for disability-related accommodations.