San José State University
College of Science/Department of Computer Science
CS 274, Topics in Web Intelligence, Section 1, Spring, 2021

Course and Contact Information

Instructor:	Dr. Teng Moh
Office Location:	Zoom
Telephone:	(408) (924-5147)
Email:	MyFirstName <dot> MyLastName <at> SJSU <dot> EDU
Office Hours:	M 17:45 to 18:45 and W 15:30 to 16:30
Class Days/Time:	MW 16:30 to 17:45
Classroom:	Zoom
Prerequisites:	CS 157A or instructor consent.

Course Description

Topics in web intelligence including finding similar items, mining data streams, link analysis, frequent itemsets, advertising on the web, recommender systems, and mining social-network graphs. Relevant applications will be covered. Significant programming project is required.

Course Learning Outcomes (CLO)

Upon successful completion of this course, students will be able to:

CLO 1 *Find similar items with local sensitive hash algorithms.*

CLO 2 *Perform data stream mining with DGIM method.*

CLO 3 *Analyze web links with PageRank algorithms.*

CLO 4 *Find frequent itemsets by a-priori algorithms.*

CLO 5 *Advertise on the Web using the BALANCE algorithm.*

CLO 6 *Build recommender systems with collaborative filtering.*

CLO 7 *Perform social-network graph mining using community detection and spectral clustering algorithms.*

CLO 8 *Implement new tools for web intelligence based on the above techniques.*

Required Texts/Readings

Textbook

- *Jure Leskovec, Anand Rajaraman, Jeff Ullman, **Mining of Massive Datasets,** Cambridge University Press, 3rd ed., 2020, ISBN: 978-1108476348*

Other Readings [Optional]

- *Ian H. Witten, Eibe Frank, Mark A. Hall and Christopher J. Pal, **Data Mining**, Morgan Kaufmann, 4th ed., 2016, ISBN: 978-0128042915*

Other equipment / material requirement

- *Wireless laptop*

Course Requirements and Assignments

SJSU classes are designed such that in order to be successful, it is expected that students will spend a minimum of forty-five hours for each unit of credit (normally three hours per unit per week), including preparing for class, participating in course activities, completing assignments, and so on. More details about student workload can be found in University Policy S16-9 at http://www.sjsu.edu/senate/docs/S16-9.pdf.

Homework is due typewritten (include source code, but not executable files) by class starting time on the due date. Each assigned problem requires a solution and an explanation (or work) detailing how you arrived at your solution. Cite any outside sources used to solve a problem. When grading an assignment, I may ask for additional information. A subset of the assigned problems will typically be graded.

Refer the course website for latest information of homework assignments.

NOTE that University policy F15-12 at http://www.sjsu.edu/senate/docs/F15-12.pdf states that "Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading."

Final Evaluation

A final paper for the semester-long programming project is due on May 19 (Wednesday).

Grading Policy

Grading information:

- *I will determine letter grades for the course, including +/- grades based on*

Percentage	***Grade***
92 and above	*A*
90 - 91	*A-*
88 - 89	*B+*

82 - 87	*B*
80 - 81	*B-*
78 - 79	*C+*
72 - 77	*C*
70 - 71	*C-*
60 - 69	*D*
59 and below	*F*

- *List of the percentage weight [or point value] assigned to various class assignments*
  - *Homework: 20%*
  - *Project Proposal: 25%*
  - *Project Presentation: 25%*
  - *Final Paper: 30%*
- ***NO** make-up exams will be given and **NO** late homework will be accepted.*

Classroom Protocol

- ***Always start your email subject with* [CS274] *to get my attention.***
- ***Wireless laptop is required**. Your laptop must remain closed (preferably in your backpack and, in any case, not on your desk) until I inform you that it is needed for a particular activity.*
- ***Cheating** will not be tolerated, but working together is encouraged*
- *Student must be respectful of the instructor and other students. For example, but not limited*
  - *Turn off cell phones*
- *To encourage participation from students, **NO** recording is allowed.*

University Policies

Per University Policy S16-9 *(http://www.sjsu.edu/senate/docs/S16-9.pdf)*, relevant information to all courses, such as academic integrity, accommodations, dropping and adding, consent for recording of class, etc. is available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/". Make sure to visit this page, review and be familiar with these university policies and resources.

CS 274, Spring 2020, Course Schedule

The schedule is subject to change with fair notice and the notice will be made available in class.

Course Schedule

Week	Topics, Readings, Assignments, Deadlines
1	Overview of Web Intelligence
2	Shingling & Min-Hashing
3	Locality-Sensitive Hashing & Sampling a Fixed-Size Sample
4	Queries Over a Sliding Window & DGIM Method
5	Web Search & PageRank
6	Block-Based Update Algorithm & Frequent Itemset Mining
7	Association Rules & A-Priori Algorithm
8	Multistage Algorithm & Performance-Based Advertising
9	BALANCE Algorithm & Content-based Recommender Systems
10	**Spring Recess**
11	Collaborative Filtering
12	Community Detection & Spectral Clustering
13	Dimensionality Reduction
14	Applications of Mining Data Streams & Link Analysis
15	Applications of Frequent Itemsets & Advertising on the Web
16	Applications of Recommendation Systems & Mining Social-Network Graphs
17	Future Direction