**Course Syllabus**

Department of Computer Science and Engineering, College of Engineering

CRN 16565/16566, Sec. 001, 3 Credits

| **Instructor** | Raiyan Abdul Baten, Ph.D. | **Term and year** | Spring 2024 |
| --- | --- | --- | --- |
| **Email** | [rbaten@usf.edu](mailto:rbaten@usf.edu) (recommended mode) | **Class days and location** | TR 3:30 P.M. - 4:45 P.M.;  CHE 303 |
| **Office Phone** | You can contact me through Microsoft Teams, but emails are recommended | **Class modality** | In-person. Possible occasional synchronous or asynchronous online instruction |
| **Office hours** | TR 11:00 AM - 12:30 PM; in-person or by appointment | **TA** | Rupal Agarwal (rupalagarwal@usf.edu) |
| **Office location** | ENG 116 | **TA office hours** | Friday, 10:30 AM-12:30 PM, via MS Teams (link: <https://teams.microsoft.com/l/meetup-join/19%3ameeting_ZTkxMzZkNGEtYWMyMS00MDM1LWI0MzgtMzU1YTA3YWQ3MDMw%40thread.v2/0?context=%7b%22Tid%22%3a%22741bf7de-e2e5-46df-8d67-82607df9deaa%22%2c%22Oid%22%3a%22cd16c4f1-6c73-4c9e-99e9-8af6de271003%22%7d>) |
| **Website** | <https://www.raiyan-ab.com> | **Credit hours** | 3 |

## **1. Welcome**

How does Facebook know who the people you went to high school with are? Who suggests what we should watch on Netflix tonight? How do memes and content go “viral” on social platforms? Learn and understand how the digital trails left behind by millions of users online can be turned into actionable knowledge for various applications.

## **2. University Course Description**

This course introduces useful techniques to model, analyze, and understand large-scale social media, focusing on social network analysis, user modeling, bot detection, and dynamic processes over social and information networks.

## **3. Prerequisites (UG section only)**

* COP 4530 Data Structures (C− minimum)
* CDA 3201 Computer Logic Design (C− minimum)

## **4. Student Learning Outcomes**

This course introduces quantitative methods for analyzing data from social media platforms. The students will demonstrate the ability to acquire, process, analyze, and visualize social media data and employ open-source toolkits such as scikit-learn or networkx for data mining. There will be a strong focus on theoretical intuition building, where the students will demonstrate a deep, intuitive familiarity with the well-documented patterns of human-human and human-content interaction in social media.

## **5. Course Objectives**

The students will be able to

1. (a) Create, (b) traverse, and (c) visualize graphs using real and synthetic social network data
2. Identify the most important people in a social network using network structure information (centrality)
3. Compute global characteristics of social networks (degree distributions, clustering coefficient, average path length) and explain their implications
4. Mine and explain how people connect with (befriend/follow) others
   1. Transitivity
   2. Homophily
   3. Mathematical models: ER, Small world, Preferential attachment
5. Identify communities in a social network
6. Mine user behavior (e.g., sentiment) on social media
7. Track information diffusion (e.g., of links and content) through social connections
8. Create basic recommendation algorithms for contents (collaborative filtering) and peers (triadic closure)

## **6. How to Succeed in this Class**

You should be familiar with basic descriptive statistics (mean, variance, etc.), probability theory (mostly discrete, e.g., chain rule, Bayes rule, etc.), and linear algebra, so please refresh them during the first weeks. A data mining or machine learning background is a plus but not required. All programming assignments will be in Python. I will list some resources to get up to speed with Python in Canvas.

There will be heavy demand for your time beyond the classroom. The nature of the topic will require you to read papers/book chapters, write reports, understand and implement algorithms, work on a research project, and present to the class.

Last but not least, although social media has implications in many disciplines (including philosophy and literature!), we will view it primarily from the lens of STEM in this course. A lot of people think that STEM topics are not for them. The opposite is true: anyone can learn this stuff. I recommend a nice essay by Susan J. Fowler (bit.ly/2ZnmzRS) about this small but important truth.

## **7. Required Texts and/or Readings and Course Materials**

* Reza Zefarani, Mohammad Ali Abbasi, Huan Liu, “Social Media Mining: An Introduction.” Cambridge University Press, 2014. ISBN: 978-1107018853.
* The [PDF of the book](https://doi-org.ezproxy.lib.usf.edu/10.1017/CBO9781139088510) is available through the USF Library (note: requires login with your USF username and password). You can buy a hard copy if you want, but it is not required.
* There will be additional required readings from the literature; see Section 13 for more information.

## **8. Course Topics**

* Graph Theory: graphs, graph traversal
* Centrality Measures: degree, eigenvectors, betweenness, etc.
* Statistical Properties of Networks: degree distributions, clustering coefficient, average path length
* Models of Networks: Random Graphs (E–R model, Small World), Preferential Attachment (B–A model)
* User behavior: Mining patterns of user connectivity (community detection) and content engagement (sentiment analysis)
* Information diffusion: Cascades, Linear Threshold Model
* Recommendation engine: Collaborative filtering, triadic closure

## **9. Schedule**

This is a tentative schedule of the course. Note that it is subject to change as the term progresses.

* **The book chapter readings are due before the start of the class.**
* **The ‘Paper Reading Assignments’, ‘Coding Assignments’, and ‘Presentation slides’ are due 12:00 pm, noon, before class on the respective due dates.**

| **Week** | **Date** | **Topics** | **Notes** |
| --- | --- | --- | --- |
| Week 1 | **1/9** | **Introduction**   * Graphs & social media | **Reading:** Chapter 1 |
| **1/11** | **Graph representations**   * Nodes and edges * Adjacency matrix/list * Weighted graphs * Directed graphs * Bipartite graphs | **Course objectives:** 1a  **Reading:** Chapter 2.1 - 2.5 (inclusive) |
| Week 2 | **1/16** | **Individual characteristics: Centrality - Part A**   * Degree centrality * Closeness centrality * Betweenness centrality * PageRank centrality | **Course objectives:** 2  **Reading:** Chapter 3.1.1-3.1.6 (inclusive) |
| **1/18** | **Individual characteristics: Centrality - Part B**   * (cont’d) | **Course objectives:** 2  **Reading:** Chapter 3.1.1-3.1.6 (inclusive) |
| Week 3 | **1/23** | ***Review 1:*** *Building and visualizing graphs, social media datasets* | **Course objectives:** 1a, 1c |
| **1/25** | **Individual characteristics**   * Transitivity * Local clustering coefficient | **Course objectives:** 4a  **Reading:** Chapter 3.2  **Coding assignment 1 due** |
| Week 4 | **1/30** | **Graph traversal**   * BFS * DFS | **Course objectives:** 1b  **Reading:** Chapter 2.6 |
| **2/1** | ***Review 2:*** *Centrality, BFS, and DFS* | **Course objectives:** 1b, 2 |
| Week 5 | **2/6** | **Mining network characteristics: Degree distributions**   * Degree distribution in a simple graph * Power law distribution | **Course objectives:** 3  **Reading:** Chapter 4.1  **Paper reading assignment 1 due:**  Barabási and Albert, “Emergence of Scaling in Random Networks” |
| **2/8** | **Network models**   * Erdos Renyi model * Barabasi-Albert model | **Course objectives:** 4c  **Reading:** Chapter 4.2, 4.4  **Coding assignment 2 due** |
| Week 6 | **2/13** | ***Review 3:*** *Network Characteristics and Models - Part A* | **Course objectives:** 3, 4c |
| **2/15** | **Mining network characteristics**   * Clustering Coefficient * Average path length * Milgram’s experiment | **Course objectives:** 3  **Reading:** Chapter 4.1, 4.3  **Paper reading assignment 2 due:**  Watts and Strogatz, “Collective dynamics of ‘small-world’ networks” |
| Week 7 | **2/20** | **Network models**   * Small world network | **Course objectives:** 4c  **Reading:** Chapter 4.3  **Coding assignment 3 due** |
| **2/22** | ***Review 4:*** *Network Characteristics and Models - Part b* | **Course objectives:** 3, 4c |
| Week 8 | **2/27** | **Community detection - Part A**   * Zachary’s Karate Club * Hierarchical clustering * Spectral clustering | **Course objectives:** 5  **Reading:** Chapter 6.1.1-6.1.3  **Project team formation** |
| **2/29** | **Community detection - Part B**   * (cont’d) | **Course objectives:** 5  **Reading:** Chapter 6.1.1-6.1.3 |
| Week 9 | **3/5** | ***Review 5:*** *Community detection* | **Course objectives:** 5 |
| **3/7** | **User behavior mining**   * Sentiment analysis | **Course objectives:** 6  **Coding assignment 4 due** |
| Week 10 | **3/12** | Spring break | |
| **3/14** | Spring break | |
| Week 11 | **3/19** | Project Proposal Presentation (Groups 1-6); slides due at noon | |
| **3/21** | Project Proposal Presentation (Groups 7-12); slides due at noon | |
| Week 12 | **3/26** | **Social influence and homophily**   * Applied modularity: Blog * Influence: Linear Threshold Model * Homophily | **Course objectives:** 4b, 6  **Reading:** Chapter 8.2, 8.3  **Paper reading assignment 3 due:** Vosoughi, Roy, and Aral, “The spread of true and false news online” |
| **3/28** | **Diffusion of information; Strength of weak ties**   * Katz Flow model * Creativity and intelligence * Early detection of outbreaks | **Course objectives:** 7  **Reading:** Chapter 7.3-7.4 (inclusive)  **Paper reading assignment 4 due:** Christakis and Fowler, “Social  Network Sensors for Early Detection of Contagious Outbreaks” |
| Week 13 | **4/2** | **Diffusion of behavior; Emotion and behavior contagion**   * Instrument variable regression * Emotion contagion * LinkedIn SoWT * Behavior contagion | **Course objectives:** 7  **Paper reading assignment 5 due:** Sinan Aral and Christos Nicolaides, "Exercise contagion in a global social network." |
| **4/4** | ***Review 6:*** *Behavior, Influence, diffusion of information* | **Course objectives:** 4b, 6, 7 |
| Week 14 | **4/9** | **Recommendation engines**   * Collaborative filtering * Triadic closure | **Course objectives:** 8  **Reading:** Chapter 9.1-9.2.2 (inclusive) |
| **4/11** | No class (tentatively, subject to change) | |
| Week 15 | **4/16** | Final Presentations (Groups 1, 2, 3); slides due at noon | |
| **4/18** | Final Presentations (Groups 4, 5, 6); slides due at noon | |
| Test free week (4/20 - 4/26) | **4/23** | Final Presentations (Groups 7, 8, 9); slides due at noon | |
| **4/25** | Final Presentations (Groups 10, 11, 12); slides due at noon | |
| Exams week (4/27 - 5/2) | **4/29** | Final Project Report due (9 am on Mon, April 29, 2024) | |

## **10. Grading**

This is a research-oriented project-based course (total of 100 points). Instead of a final exam, each student will do 5 pop quizzes, 5 paper reading assignments, 4 coding assignments, and a group research project.

**Pop quiz:** The pop quizzes will be taken during class via pencil-and-paper. **Expect a pop quiz every class since these quizzes will be taken without explicit announcements.** There will be a total of 5 quizzes. The instructor may choose to add bonus quizzes later.

**Paper Reading Assignments:** There will be 5 required papers to read. Each student should read the assigned paper and submit a short critique (100–200 words) online in Canvas before 12:00 pm (noon) on the respective due dates. These reviews should not be simple summaries but discuss positive aspects of the paper and limitations (examples: https://nlpers.blogspot.com/2016/08/some-papers-i-liked-at-acl-2016.html) or suggestions for how the work could be improved or extended. Submissions that simply state a summary of the paper will be assessed a grade of ‘incomplete.’

**Coding Assignments:** There will be 4 coding assignments. Each assignment will consist of auto-graded coding problems. Each student should submit the solutions to the problems in Canvas before 12:00 pm (noon) on respective due dates. You will have to code in Python for your assignments due to auto-grading.

Solutions to your auto-graded coding problems must be self-sufficient and not dependent on other data mining / social network analysis packages or code, such as NetworkX or scikit-learn (unless explicitly specified). You may use packages for display graphics or mathematics packages, such as routines for data processing, linear algebra, statistics, and optimization — for example, pandas, numpy, or scipy (but not the Compressed Sparse Graph Routines ). You may also use Tweepy or other networking libraries (e.g., ScraPy, requests) to collect data from Twitter.

**Final Project:** For your final project, you may use other packages or code and a language other than Python. Detailed instructions for the Final Project will be announced later.

## **10a. Grading categories and weights**

* Attendance (individual, 5% of final grade),
* 5× Pop quizzes (individual, 10% of final grade),
* 5× Paper reading assignments (individual, 5% of final grade),
* 4× Coding assignments (individual, 40% of final grade divided as 10% each),
* Group project (40% of final grade, divided as 30% common + 10% individual; groups of 5 max):
  + Project Proposal Presentation, 10% (entirely common grade)
  + Final Presentation, 15% (= 10% common grade + 5% individual grade)
  + Written Report, 15% (= 10% common grade + 5% individual grade)

## **10b. Grading scale**

Grading Scale (%)

94 – 100 A 74 – 76.999 C

90 – 93.999 A- 70 – 73.999 C-

87 – 89.999 B+ 67 – 69.999 D+

84 – 86.999 B 64 – 66.999 D

80 – 83.999 B- 60 – 63.999 D-

77 – 79.999 C+ 0 – 59.999 F

## **10c. Late work policy**

There will be no make-up for reading assignments, coding assignments, and pop quizzes.

At the beginning of the semester, each student has six (6) extension tokens, each equivalent to a 24-hour extension on the due date of a reading assignment or a coding assignment. To use a token, please contact the TA or the instructor. No reasons need to be provided. Tokens cannot be used on pop quizzes, group projects, or for attendance. No additional tokens will be provided after the 6 tokens are used.

Please e-mail your reading/coding assignments to the instructor if there are any technical issues with online submission. Reading/coding assignments turned in late will be assessed a 25% penalty of the earned grade each late day.

## **10d. Regrade policy**

If you believe an error has been made in grading your work, you may resubmit it for a regrade — submit a detailed explanation of which problems you think we marked incorrectly and why. Because we will examine your entire submission in detail, your grade can go up or down due to a regrade request.

## **10e. Group Work Policy**

Everyone must take part in a group project. The project proposal presentation will be graded as a team (everyone in the project gets the same score). The final project presentation (15%) and the final written report (15%) will both comprise team (10%) and individual (5%) components. For the team component, everyone in the team will get the same score based on the assessment of the project. The instructor will determine the individual component based on the declaration of work division in the project report and performance during the final project presentation. Once formed, groups cannot be altered or switched, except for reasons of extended absence due to medical reasons.

## **10f. Midterm grade (CAP 4773 only)**

There is no midterm exam in this course. However, per USF System Policy 10-504, a midterm grade will be made available to you in OASIS. It will reflect the attendance rate and any graded assignments up to that point.

## **11. Course Policies: Technology and Media**

**Canvas:** This course will use USF’s learning management system, Canvas. **You are responsible for reading and being aware of ALL ANNOUNCEMENTS in Canvas and EMAIL sent to your official @mail.usf.edu / @usf.edu email accounts.** If you need help learning to perform various tasks related to this course, please consult the Canvas help guides. You may also contact USF’s IT department at (813) 974-1222 or help@usf.edu.

**Internet Access:** You are expected to have and maintain Internet access and a device by which you can access the Internet and web content posted on Canvas.

**Laptop Usage:** The use of a laptop during the lecture is allowed only for capturing notes and looking up course/lecture-related materials. Use of a laptop for other purposes is prohibited.

**Classroom Devices/Student Recording**: The use of tape recorders or other recording devices is NOT allowed in the class.

**You do not have the right to sell notes or tapes of lectures generated from this class.**

**Phone Usage:** Using a phone during class is prohibited, including texting or surfing the Internet, unless explicitly permitted by the instructor. Students may not take photos/video/audio recordings of the class lectures. Only pictures of the slides/whiteboard notes are allowed (the slides will be uploaded to Canvas anyway).

## **12. Standard University Policies**

## **USF Core Syllabus Policies**

USF has a set of central policies related to student recording class sessions, academic integrity and grievances, student accessibility services, academic disruption, religious observances, academic continuity, food insecurity, and sexual harassment that apply to all courses at USF. Be sure to review these online: usf.edu/provost/faculty-success/resources-policies-forms/core-syllabus-policy-statements.aspx

## **Attendance policy**

The students are expected to attend classes. Excused absences can be accommodated (e.g., related to university-sponsored events, jury duty, and other cases), but the student needs to inform the instructor ahead of time (when possible). The exact form of accommodation will be decided on a case-by-case basis. There is no makeup for any in-class pop quizzes unless the Instructor determines a special need. See Section 10c regarding extension tokens for paper reading assignments and coding assignments. Documented excused absences for any group project activity (proposal presentation, final presentation, final report) may be allowed by making arrangements ahead of time (when possible) or by providing a reasonable amount of time to make up for the missed work.

## **TurnitIn / MOSS**

In this course, we may use Turnitin (http://turnitin.com/), an automated system instructors use to quickly and easily compare each student’s assignment with billions of websites and an enormous database of student papers that grows with each submission. After the assignment is processed, the TA and I will receive a report showing if and how another author’s work was used in the assignment. We may also use the MOSS code checker (https://theory.stanford.edu/~aiken/moss/) to check for code plagiarism.

## **13. List of Readings**

[1] Albert-László Barabási and Réka Albert. “Emergence of Scaling in Random Networks”. In: *Science* 286.5439 (1999), pp. 509–512. ISSN: 0036-8075.

[2] Nicholas A. Christakis and James H. Fowler. “Social Network Sensors for Early Detection of Contagious Outbreaks”. In: *PLoS One* 5.9 (Sept. 2010), pp. 1–8.

[3] Soroush Vosoughi, Deb Roy, and Sinan Aral. “The spread of true and false news online”. In: *Science* 359.6380 (2018), pp. 1146–1151

[4] Duncan J. Watts and Steven H. Strogatz. “Collective dynamics of ‘small-world’ networks”. In: *Nature* 393.6684 (June 1998), pp. 440–442. ISSN: 1476-4687.

[5] Sinan Aral and Christos Nicolaides. "Exercise contagion in a global social network." *Nature Communications* 8, no. 1 (2017): 14753.