CS9.435 Computational Social Science
Mid-Semester Exam

March 02, 2022
Max marks: 31
Allotted Time: 48 Hours

Instructions

- Please create a single **typed** PDF answering the following questions and submit it to **Mid Semester** assignment on MS Teams. Name the file **<roll-number>_midsem.pdf**.
- The duration of the mid-sem is 48 hours. Please submit your PDFs on Teams by then.
- All submissions will be tested for plagiarism, and if found, the institute's policies will be followed.
- State any assumptions you make while answering the questions.
- For questions that require you to draw diagrams/figures, you can either
    - Create them on draw.io and paste them in your typed document.
    - OR Hand-draw them and merge a scanned PDF of the same with your typed document

Questions

PART A [15 marks]
Attempt all questions in this part.

Q1 [15 marks, 900-1400 words]

Read and summarize the following paper:
Cheng, Justin, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 'How Community Feedback Shapes User Behavior'. In Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014.

a) [8 marks, 500-600 words]
Write a detailed summary of the paper's objectives, research questions/hypotheses, methodology, and results.
b) [4 marks, 300-500 words]
Express your opinions on the paper in terms of its strengths and weaknesses, what it should have done differently, possible extensions, and its utility.
c) [3 marks, 100-300 words]
The paper follows a common paradigm of verifying existing psychological theories (in this case, the operant conditioning framework) using large-scale data collected from social media and other sources. What do you think are the benefits and possible pitfalls of such analysis?

PART B [16 marks]

Attempt any 2 out of the 3 questions in this part. Please indicate the correct question number. In the case more than 2 questions are attempted, the first two would be evaluated.

Q2 [8 marks, 400-700 words]

The paper by Kramer et al. on massive-scale emotional contagion through social networks has come under ethical criticism for its experimental design and the impact it could have on its subjects. Suggest at least 3 detailed measures to redesign the study such that it is less invasive and has a lesser negative impact on the participants of the experiment. Elaborate on the differences between your proposed alternatives and the paper's approach.

Q3 [8 marks, 400-700 words]

Collecting redundant data is a useful way to assess the quality and consistency of distributed data collection. Read this paper by Windt et al. which develops and tests a system to collect reports of conflict events.

a) [3 marks, 100-200 words]
   How does their design ensure redundancy?
b) [3 marks, 200-300 words]
   They offer several approaches to validate the collected data. Summarise them. Which one do you find the most convincing?
c) [2 marks, 100-200 words]
   Propose a new way of validating the collected data. Try to increase confidence in the data while keeping the collection cost-effective and ethical.

Q4 [8 marks, 400-700 words]

Amazon Mechanical Turk is a common crowdsourcing platform employed by many computational scientists for their research. Researchers use MTurk to access thousands of on-demand workers for different tasks ranging from annotations to data processing etc. Imagine the experience of an MTurker and then assess the design, quality, and ethics of human computation projects.

a) [4 marks, 200-400 words]
   Critically evaluate human computation experiments on overall research quality, experiment design, and ethical grounds.

b) [4 marks, 200-300 words]
   Describe in detail how crowdsourcing human computation efforts helps in improving data quality but does not help in removing bias.