CS 8803 Social Computing: Introduction

Munmun De Choudhury
munmund@gatech.edu
Week 1 | August 18, 2014
Part I: Course Structure and Information
Learning Objectives

• Social computing is one of the biggest forces on the internet

• **Goal: How and why social computing works?**
  • What's the right way to design social computing systems? What sense can you make out of all the data people leave behind? What's still out there to infer, understand, and build?

• **At the completion of this course, you’ll be able to:**
  • Understand important features of social computing
  • Design and prototype new social computing systems
  • Analyze data left behind in social media
  • Understand the research issues in this field
Grading

- Responses to Class Readings (on Piazza) - 10%
- Assignment I - 10%
- Assignment II - 10%
- Assignment III - 10%
- Data Project Phase I Report (due at midterm) - 20%
- In-class Presentation of Data Project Phase I - 5%
- Data Project Phase II Report (due at final) - 20%
- In-class Presentation of Data Project Phase II - 5%
- Class Participation – 10%
Required Skills

• Technical: any object-oriented/scripting language like Python, Perl, C#; some frontend development skills/web programming skills (ajax, javascript, php) if your project is about building a social tool

• Statistics/data mining: preliminary knowledge of working with some data and using some analytical software

• Reading/writing: significant weightage on weekly class readings, mid-term and final term project report
No required text, only suggestions

- **SNA:**
  - Networks, Crowds, and Markets, by David Easley and Jon Kleinberg
  - Six Degrees, by Duncan Watts

- **Social Science-y:**
  - On Individuality and Social Forms, by Georg Simmel
  - Networked, by Barry Wellman

- **Machine Learning:**
  - Machine Learning for Hackers, by Drew Conway and John Myles White
  - Natural Language Processing with Python, by Steven Bird, Ewan Klein, and Edward Loper
  - Pattern Classification, by Richard Duda, Peter Hart, and David Stork

- **Pop Statistics:**
  - The Signal and the Noise, by Nate Silver

- **Writing:**
  - Writing for Social Scientists, by Howard Becker
Participation in Class Readings

- Readings are available on class website
  - Check the class webpage before each class, as readings are subject to change

- Write short blurbs on Piazza (under “readings”) about the readings assigned for a particular class
  - Blurbs can range from 200-500 words in length

- Blurbs should focus on the following, but not limited to:
  - The novelty in the paper you read i.e., what was the contribution
  - Ways the work in the paper could improve
  - Why the particular method/data used in the paper was appropriate
Piazza

- Sign up with your GT information at this link [IMPORTANT]:
piazza.com/gatech/fall2014/cs8803specialtopics

- Class link:
piazza.com/gatech/fall2014/cs8803specialtopics/home
Assignment I

- Due: September 10, 2014
- **Topic:** Propose the design and implement a tool that would give real-time notifications of “trending topics” based on your Facebook/Twitter friends

- What to hand in?
  - 3 page report and 1 page extra for a mockup of what the tool will look like
  - Submission on T-Square
Assignment II

- Due: October 8, 2014
- **Topic**: Identify textual attributes from an SMS spam dataset (provided by instructor) using the NLTK library in Python, distinguishing between features of spam messages and non-spam messages (some statistical analysis required)

- **What to hand in?**
  - 3 page report with summary of findings
  - Submission on T-Square
Assignment III

• Due: November 12, 2014
• *Topic*: In a dataset of user friendships and checkins (provided by instructor), examine whether friends checkin at similar places and if so, what network measures indicate high checkin correlation (some statistical analysis required)

• What to hand in?
  • 3 page report with summary of findings
  • Submission on T-Square
Term Project

- Project idea write-up due: September 30, 2014
- Mid-term presentations due: October 15, 2014
- Mid-term report due: October 20, 2014
- Final presentations due: December 3, 2014
- Final project due: December 5, 2014

Goals:

- Group project: 3-4 people; total about 10 groups in the class
- You are free to pick your group
- Need to discuss your project idea with instructor early on in the course. Instructor will provide some sample projects in the class
- Presentations (both midterm and final) will be total 8 minutes, with 6 minutes talking about task accomplished, and rest for Q&A and switching to next presenter
Finding Data

• Based on the project, you may need data.

• Suggestions:
  • Your own social media data. E.g., you might consider downloading all of your email or IM logs. Or, you could download all your Facebook data.

• Hillary Mason's research dataset page: https://bitly.com/bundles/hmason/1

• Jure Leskovec’s SNAP page: https://snap.stanford.edu/data/

• You may also decide to build a social tool in which case you will have to use different design principles and show the entire design process to get credit.
What to hand in?

- Midterm:
  - If building a tool: design process, mockup, and an early prototype (i.e., demo)
  - If analyzing data: data collection method/key properties of the data, plan for analysis
  - Report length: 4-5 pages, single column format submitted through T-Square

- Final:
  - If building a tool: working prototype, and proposal of a method of its evaluation
  - If analyzing data: choice of the analysis technique (text vs. network analysis), key findings from the analysis (e.g., charts and tables)
  - Report length: 8-10 pages, single column format (can include content from midterm for continuity purposes) submitted through T-Square

- Clearly articulate in an extra page individual contribution
- Typically you will not need to submit the code, unless some exception arises
- Okay to use open source tools (in fact encouraged)
Late Policy

• Reading responses are due at 11:59pm on the day before the relevant class meeting.

• Assignments are due one hour before class (i.e., 3:35pm) on the date listed for that assignment.

• Any work (reading response, assignment, or project) submitted more than 15 minutes after the due time will be assessed a 10% penalty.

• Each additional 24 hours of lateness will result in an additional 10% being taken off the grade for that assignment. After 5 days, the assignment will not be accepted and a grade of 0% will be entered.
English as Second Language

• If English is not your first language, you may request to not be graded on your writing for a particular individual assignment.
  • This means you won't be penalized for bad writing, but you also won't get credit for good writing. To take advantage of this option, you must mark "ESL" (English as a Second Language) on the first page of your assignment/paper.

• This option is not available for the term project as it is a group assignment.
Academic Integrity

• This class abides by the Georgia Tech Honor Code.
• All assigned work is expected to be individual, except where explicitly indicated otherwise.
• You are encouraged to discuss the assignments with your classmates; however, what you hand in should be your own work.
  • Okay to use open-source software (no need to reinvent the wheel), however do acknowledge!
  • Copying/reusing code from your classmates and friends are not allowed; strict action will be taken if similarities are discovered
  • Copying (textual) content for your assignments and project from other published work (without citing them) is also not allowed, and is considered plagiarism
Help and Resources

- Office hours: 11am – 12 noon Thursday, or by appointment
- Location: TSRB 236

- Teaching Assistant: Joe Gonzales
- Office hours: TBA
- Location: TBA
- Email: jgonzales8@gatech.edu

- Class website (including readings): http://www.munmund.net/CS8803.html

- Email announcements will be made over the course page on T-Square
Overrides

• If you couldn’t find room to register, enter your name and GT email in the sign up sheet with me before your leave the class today (only your name, do not include your friend’s name!)
  • If you have emailed me earlier about an override, indicate that in the sign up sheet

• Possibly can’t give overrides to everyone, sorry!
  • Room size constraints
  • I’ll use an internal selection process
Part II: Introductions
name + program
tech + design + stats background
what you want to learn from the class
Part II: Defining “Social Computing”
Defining Social Computing by Tom Erikson, IBM Watson Labs
Topics to be covered

• **Key characteristics**
  • SNS, social media overview; design of social systems
  • Discourse and dialogue; social capital

• **Methods**
  • Statistics, data mining review
  • Text analysis; network analysis

• **Key issues of social computing**
  • Credibility; trust; polarization; privacy

• **Uses of social computing systems**
  • Personality; trends and forecasting; event analytics; social news; social search and Q&A; location and mobility