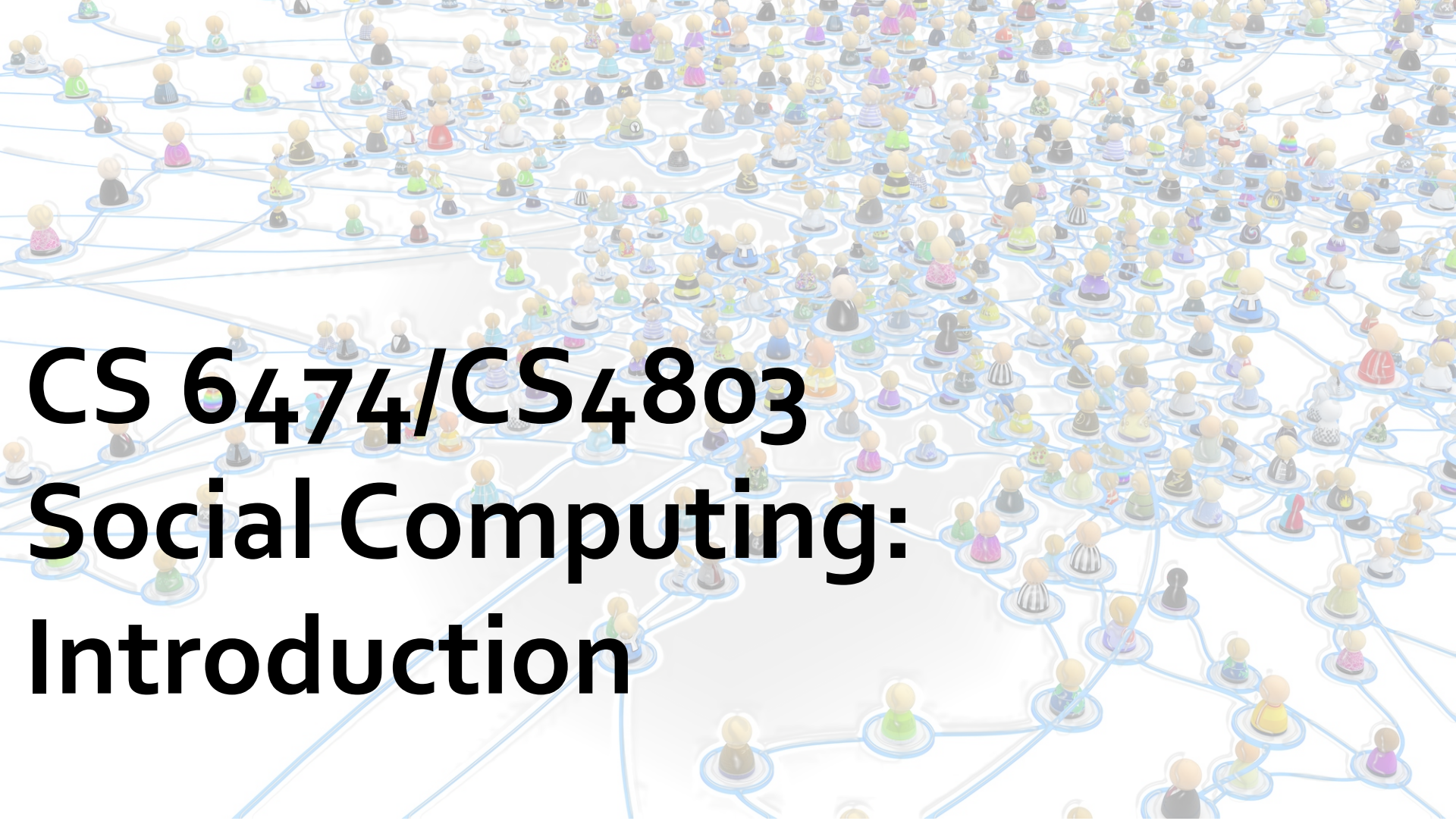


CS 6474/CS4803 Social Computing: Introduction

Munmun De Choudhury

munmund@gatech.edu

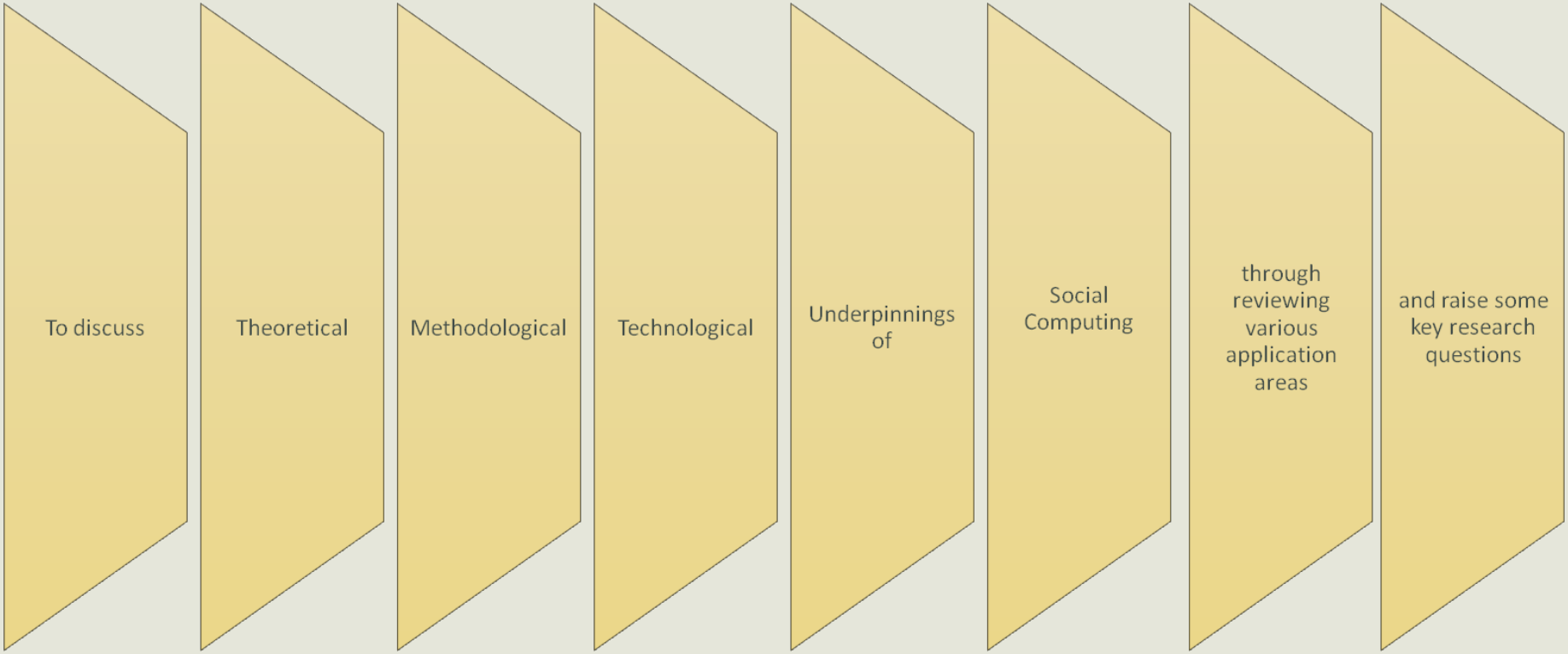
Week 1 | August 20, 2018

Part I: Course Structure and Information

Learning Objectives

- Course Website: http://www.munmund.net/CS6474_Fall2018.html

- **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?
  - How has social computing impacted the world? What benefits and challenges has it presented?
- 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 (theoretical and practical) in this field


To discuss
Theoretical
Methodological
Technological
Underpinnings of
Social Computing
through reviewing various application areas
and raise some key research questions

Topics to be covered

- Sociological Foundations
- Social Computing Theories
  - Public displays and performance, identity, self-disclosure, social capital, social influence
- Analysis: Language
  - NLP and text analytics
- Social Computing Constructs and Issues
  - Credibility and trust, polarization, reputation, moderation
- Benefits/Applications of Social Computing
  - Politics, crisis, social movements and activism, deviance, prediction/forecasting
- Challenges of Social Computing Systems
  - Privacy, ethics

- Methods
  - Statistics, data mining review, machine learning and NLP applications

Suggested books

- 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

Grading

- Responses to Class Readings - 20% (Best 16 out of 19)
- Assignment I - 10%
- Assignment II - 10%
- Assignment III - 10%
- Class Participation - 10%
- Term Project - 40%
  - Project Proposal - 5%
  - Midterm Project Presentation - 5%
  - Midterm/Milestone Report - 10%
  - Final Project Presentation/Demo - 5%
  - Final Report - 15%

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/machine learning:* preliminary knowledge of working with some data and using some analytical software (e.g., R)

- *Reading/writing:* approximately two papers assigned for each class (four in a week); significant weightage on weekly class readings, mid-term and final term project report

Important notes about skills

- No programming will be taught – you are required to have fairly good working knowledge of data analytics

- Some review of data analytics (two classes) and a working group session (one other class)
  - No detailed coverage of the methods
  - Need to have working knowledge of the data analytic methods (e.g., basic NLP, machine learning – supervised and unsupervised)

Course Materials/Logistics

- Links to papers assigned to classes will be available on the course website.
  - To access ACL Digital library links to papers, use GT VPN.
  - Paper assignments are subject to change, so always check the online schedule before you read and write your reflections.
- No official text books.
- Use of open source libraries encouraged.
- Assignments will be released 3 weeks ahead, on T-Square.
- Lecture slides of each class will be made available on the course website within a week's time.
- Classes will involve time to time in-class activities – encouraged to bring laptops, but strictly restricted to classroom needs and use.

Participation in Class Readings

- Write short blurbs on Piazza (under "reflections") about the readings assigned for a particular class
  - Blurbs can range from 300-600 words in length

- Blurbs should focus on the following, but not limited to:
  - What is the problem the paper is studying? Why is it important?
  - How does the paper relate to the topics being covered in the class?
  - What is the novelty of the paper?
  - Ways the work in the paper could improve
  - Why the particular method/data used in the paper was appropriate
- Sample reflections will be emailed by Wednesday
- Total 19 reflections
  - **Starts from next week**

Piazza

- Sign up with your GT information at this link [IMPORTANT – needs to be completed by Wednesday]:

  piazza.com/gatech/fall2018/cs6474

- Class link:

  piazza.com/gatech/fall2018/cs6474/home OR https://piazza.com/class/jl1mkxmh83m746

Assignment I

- **Due: September 12, 2018**
- Questions will focus on preliminary material covered in the class.
- Questions will test your basic understanding and thinking process of social computing systems/experience of using social computing platforms
  - Programming not required

- What to hand in?
  - 4 page report of your answers (single column, double spacing, 11 point font)
  - Submission on Canvas

Assignment II

- **Due: October 10, 2018**
- Questions will focus on the topics of the reading materials covered until the point the assignment is released.
- ***Design focused questions*** – questions will ask you to develop a mockup of a web application, browser extension etc.
  - Programming not required

- What to hand in?
  - 5-8 page report + screenshots of the mockup (single column, double spacing, 11 point font)
  - Submission on Canvas

Assignment III

- **Due: November 14, 2018**
- Questions will focus on the topics of the reading materials covered until the point the assignment is released.
- *Data analytic questions* –Questions that would test your thinking and formulation of applying data analytic approaches to a social computing question
  - Programming required

- What to hand in?
  - 5-8 page report (single column, double spacing, 11 point font) + code in a zipped folder
  - Submission on Canvas

Term Project

- Project proposal due: September 26, 2018
- Mid-term presentations due: October 29, 2018
- Mid-term/Milestone report due: October 31, 2018
- Final presentations due: November 28 + December 3, 2018
- Final project report due: December 10, 2018
- Goals:
  - Group project: 3-4 people
  - You are free to pick your group – use Piazza to know and find like-interested classmates
  - Need to discuss your project idea with instructor/TA early on in the course, before proposals are due
  - Presentations (both midterm and final) will need to focus on the project goals, challenges faced, and task accomplished; a few minutes for Q&A

Term Project

- Individual assessment – clearly articulate individual goals and contributions in the project proposal
  - In midterm and final reports, revisit the above list to indicate what you have done
- Peer assessment – indicate how each of your groupmate accomplished what they were supposed to do
  - Only required in the final report
  - Anonymous to teammates

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.
  - Crawl data from the web, with or without using APIs of the respective platforms.
- Jure Leskovec's SNAP page: https://snap.stanford.edu/data/

Late Policy

- Reading reflections are due *at 11:59pm* on the day before the relevant class meeting.
- Assignments are due *at 11:59pm* on the date listed for that assignment.
- Term project reports (midterm, final) are due at 11:59pm on the date listed.
- Some work (only assignments, or the project deliverables, but not the presentations) submitted more than 15 minutes after the due time will be assessed a 25% penalty.
  - Each additional 24 hours of lateness will result in an additional 25% being taken off the grade for that assignment. After 2 days, the assignment will not be accepted and a grade of 0% will be entered.
- **No extensions for reading reflections.**

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: By appointment
- Location: TSRB 341

- Teaching Assistant: Koustuv Saha
- Office hours: By appointment
- Location: TSRB 341A
- Email: koustuv.saha@gatech.edu

- Email announcements will be made over the course page on Canvas or on Piazza

If you need to reach me or the TA…

- Questions should be directed via email to me or TA for fastest response
- For questions/concerns related to the assignments or project deliverables, reach us at least 2 days (48 hours) before the due date.

- Questions within 2 days (48 hours) of the due date should not be expected to be answered on time.

Part II: Defining “Social Computing” / Background

Quoting Wikipedia:
"*Social computing is an area of computer science that is concerned with the intersection of social behavior and computational systems. It is based on creating or recreating social conventions and social contexts through the use of software and technology*."

Why Social Computing?

Interact

- Expressions
- Gestures
- Spoken Word
- Written Word

Sensitive to the people around

Humans are Social

Make decisions shaped by social context

- Choosing a restaurant
- Crossing the street

Doing what others do and following what others say

Theoretical and Infrastructure Basis of Social Computing

- Social Computing is a cross-disciplinary research and application field with theoretical foundations including both ***computational*** and ***social sciences***
- To support social interaction and communication, it relies on:
    - Communication
    - Human Computer Interaction
    - Sociological, Psychological Economic, and anthropological theories
    - Social network analysis

Background: Two Distinct foci of Social Computing

- **Technological issues:**

  Interfaces and user acceptance, social effects around group collaboration and online communication

- **The use of computational techniques**

  Principally simulation techniques, to facilitate the study of society and to test out policies before they were employed in real-world organizational or political situations

Major Application Areas of Social Computing

Social computing applications are driven by the needs to:

- Develop better social software to facilitate interaction and communication among groups of people (or between people and computing devices),
- Computerize aspects of human society, and
- Forecast the effects of changing technologies and policies on social and cultural behavior

Class Activity

A) An example of a social computing system (that exists online)

Why?

B) An example of a non-social computing system (that exists online)

Why?

Social Computing Tools

BLOG

WIKI

Social Networks

RSS

Social Bookmarking

VOIP

Others

- Internet Forums
- Multimedia Sharing
- Virtual Reality

people creating
(blogs, user-generated content and podcasts)

people connecting
(social networks and virtual worlds)

people collaborating
(wikis and open source)

people reacting
(to each other: forums, ratings and reviews)

people organizing content
(tags)

people accelerating consumption
(RSS and widgets)

Source: Forrester Research (www.forresterresearch.com)

A Brief Historical Analogy

Background

- The idea of social computing can be traced back to the 1940s in Vannevar Bush's seminal 1945 Atlantic Monthly paper "As We May Think."
- In the paper, he conceived
- a memory and communication device called a memex
- his far reaching ideas before anyone has researched on them include:
- Augmentation
- Groupware and Computer-supported collaborative work

"As We May Think" predicted (to some extent) many kinds of technology invented after its publication, including hypertext, personal computers, the Internet, the World Wide Web, speech recognition, and online encyclopedias such as Wikipedia: "*Wholly new forms of encyclopedias will appear, ready-made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified*."

Background

- 1960s: J.C.R. Licklider headed the Advanced Research Projects Agency (ARPA) and cowrote a paper on “The Computer as a Communication Device” with Robert Taylor
- In this paper, Licklider and Taylor outlined methods of computer-aided group collaboration
- ARPA ultimately led to ARPANET, the predecessor to Internet.
- Meanwhile, Douglas Englebart’s lab at SRI created the first hypermedia online system, NLS (oNLine System).
- 1970s: The first collaborative software, EIES (Electronic Information Exchange System)
- 1980s: Groupware appeared

Early Efforts of Social Computing

- IBM first developed a **multiparty chat environment**, **Babble**, in 1997
- Babble and its Web-based successor, **Loops**, can support synchronous and asynchronous textual conversation among small to medium-sized workgroups
- Microsoft's **Wallop** project provides a tool that enables users to author lightweight content online and build conversations in the **context of their social networks**.
- In addition to Microsoft and IBM, many research labs and companies, including Intel,
- FXPAL, HP, PARC, Mitsubishi, MITRE,AT&T, Nokia, NASA, and Google, actively conduct
- social computing research

Early Efforts of Social Computing

Since 2004, Microsoft Research has hosted annual social computing symposiums to help promote this area of research.

IBM also hosted services science symposiums during the same time period, part of which has focused on **compartmentalizing social computing modules as services**

Purpose:

Social computing and online communities are ***changing the fundamental way people share information and communicate***

Part IV: Introductions

- name + program
- academic background
- closest social computing project, if any
- what you want to learn from the class

Handout – a survey for you to fill about the class