SOCIAL COMPUTING CSE 398/498

Class Meetings	TR 9:50 a.m. - 11:05 a.m.
Office Hours	ericpsb.youcanbook.me
Location	Building C, Room 210
Semester	Fall 2019
Professor	Eric P. S. Baumer {ericpsb@lehigh.edu}

> "*To term some media 'social' implies that there are other media that are perhaps anti-social, or even not social at all—asocial. It also invites comparisons between media based on how social each medium is. But each medium is social in its own unique way and invites particular social behaviors, its own form of sociality.*" – Zizi Papacharissi

With that caveat in mind, this class interrogates social computing from two angles. First, it covers methods for exploring how social phenomena manifest via computing technologies. Second, it examines how notions of social theory are codified in computational analytic methods. The computational techniques involved include machine learning, natural language processing, and other data science style methods. The theories involved stem from sociology, linguistics, and other socio-behavioral sciences.

Thus, we define our terms as follows:

- "Computing" refers both to the technologies that humans use to interact with one another, and to the methods we will use to analyze those interactions.
- "Social" refers both to the phenomena surrounding human interactions, and to the theoretical apparatuses that we use to make sense of those phenomena.

The course examines the interplay between these two, between the social and the computational. It is neither strictly a sociology course, nor is it strictly a computing course. By the end of the course, you should feel equipped to select the proper tools for analyzing data related to human social interactions, where "tools" includes both the computational techniques involve and the underlying theoretical frameworks.

COURSE PHILOSOPHY

This class takes a hands-on approach. You will have a chance to gain experience in conducting a computational analysis of data about social interactions and other phenomena. Almost every

class meeting will involve significant interaction.

At the same time, this class will involve a significant amount of reading. This reading will span different literatures across computing, sociology, and related disciplines. Hopefully, every student will, at some point in the semester, read something with which they are not entirely comfortable. In the spirit of Haraway, though, we will be "staying with the trouble," using those moments of slight discomfort to spur our development.

Activities

The work completed for this class will take a variety of forms.

Readings and Responses

Many days on the syllabus list one or more readings. The readings for a given day are to be completed *before class* on that day. Furthermore, you should come to class ready to discuss the readings. That means that you must have taken notes and developed questions based on the readings in advance.

Whether or not students ared require to submit these responses in a written form in advance will depend on the quality of discussion and interaction that emerges during class meetings.

Discussion and Leading

There will be very little lecturing on the material in this class. Instead, we will engage in discussions about the readings. Again, this is where your notes and questions will come in handy. That said, there will always be an appointed discussion leader who should guide us through the readings.

At the beginning of the semester, the instructor will lead all class sessions. After a few weeks of demonstrating, students will take turns leading discussion. Whether this is done individually or in groups depends on a student's course number (398 vs. 498) and the total number of students enrolled in the course. Each day should begin with some sort of hands-on activity related to the reading, as led by the discussion leader for that day. This activity should then be followed by questions and prompts related directly to the readings. Again, the instructor will demonstrate this general approach during the first few weeks of class.

Technical Days

As mentioned above, social computing is both highly theoretical and highly technical. In addition to in-class discussions of the readings, we will also conduct in-class explorations of computational techniques. Most of these will take advantage of the Jupyter Notebook environment provided by Google Colaboratory.

Such days on the syllabus will often still involve readings, which still need to be completed before class. However, these readings will be of a more technical nature, helping to provide you scaffolding for the in-class activities.

Semester Project

The majority of your grade for this class will consist of the completion of a significant project applying computational methods to understanding social phenomena. The course is structured to provide exposure to a variety of topics and methods within the first several weeks. Using these as inspiration, students will formulate a plan, including: the research question the project will explore, the data set that it will analyze, the computational methods that will be applied, and hypotheses or other expectations about the outcome(s).

Graduate students (those enrolled in 498) will complete this project individually. Undergraduate students (those enrolled in 398) may choose to complete this project as a group or individually. Thus students completing their project as a group will also complete at least one peer evaluation of their group members to ensure roughly equitable distribution of effort within the group.

Several checkpoints will provide the opportunity to receive feedback and improve the project along the way. Near the middle of the semester, students will have the opportunity to "pitch" the above mentioned plan for their projects. Students will receive feedback on this pitch, both from the instructor and from fellow class members. The time given for each student will be roughly $75 / n - 1$ minutes, where n is the number of students enrolled in the course. Later in the semester, students will present an oral, in-class progress report with preliminary results. This will, again, provide the opportunity for intermediate feedback from both the instructor and fellow class members. Neither the pitch nor the progress update need make use of presentation software (e.g., PowerPoint), though they may if desired. However, students are encouraged to make use of a Notebook in their progress report.

During the last week of the semester, all students will make a formal presentation on their projects during class. This presentation will be analogous to a research conference presentation, meaning that some form of presentation software should be used. The time given for each student will be roughly $2 * 75 / n - 2$ minutes, where n is the number of students enrolled in the course. After their presentation, students will upload a copy of their presentation materials (e.g., a slide deck) as a .pdf file on CourseSite. Presentation order will be randomized and will likely span multiple days.

The project's final deliverable will consist of two main components: technical and written. The technical component should be implemented using a Jupyter Notebook. The instructor will provide an introduction to the use of this tool for data analysis. While course instruction will be conducted in Python, students may use a different, Notebook-compatible language of their choosing (see https://github.com/jupyter/jupyter/wiki/Jupyter-kernels for a full list of available options). Students may also make use of any and all libraries or packages they deem relevant. Students should note, however, that use of tools other than those covered in class reduces the chance that the instructor will be able to provide meaningful support should they encounter implementation issues. This technical portion of the final deliverable will be submitted in the

form of a Jupyter Notebook.

The written portion of the deliverable should comprise a publication-quality research report. This report will be written using LaTeX, though instruction on LaTeX will not be provided in class. Graduate students should already be familiar with this writing environment. Undergraduate students may need to consult one of the numerous sources freely available online to familiarize themselves. Lehigh also provides institutional access to Overleaf, an online LaTeX authoring tool with a variety of useful features. The instructor will provide a set of LaTeX template files. The final written report will be submitted as a .pdf via CourseSite.

Both of these final deliverables, the technical and the written, will be due near the beginning of the exam period, roughly one week after the last class meeting. This timing will provide the opportunity to incorporate any feedback received during the in-class project presentations into the final written report.

Students who complete exemplary projects should consider submitting their work to be reviewed for potential publication in an academic or scholarly venue. Examples of such venues include the ACM's Conference on Computer Supported Cooperative Work and Social Computing (CSCW), the AAAI International Conference on Weblogs and Social Media (ICWSM), or the ACM's Transactions on Social Computing (TSC). Interested students should discuss this possibility with the instructor.

ATTENDANCE

Attendance is **mandatory**. Class time will be spent engaging in discussions of the readings, hands-on implementation activities, and other exercises. As such, in-class activities are integral, both to your own and to others' learning, as well as fulfill a non-negligible portion of your grade.

Sometimes, though, life happens. In such instances, you will receive **two personal days** to use at your discretion during the semester. There is no need to notify the instructor, and no excuse or justification need be given. Your in-class participation grade that day will not count toward your total. If you miss additional days, those will be counted as a zero toward your in-class participation grade.

GRADING

You will receive a grade based on the following break down.

10%	Attendance & Participation
10%	Leading Class Discussion(s)
5%	Semester Project – Pitch
5%	Semester Project – Progress Update
10%	Semester Project – Final Presentation

60% Semester Project – Final Written Report and Technical Analysis

Grading Scale

97% – 100% A+	77% – 79.9% C+
93% – 96.9% A	73% – 76.9% C
90% – 92.9% A-	70% – 72.9% C-
87% – 89.9% B+	67% – 69.9% D+
83% – 86.9% B	63% – 66.9% D
80% – 82.9% B-	60% – 62.9% D-
	< 60% F

Policies

Technology in Class

While in class, your attention should be on class. Please silence, disable, or turn off any device that makes noise.

Use of computational technology during class will vary, depending on the in-class activities. On some days, you will be asked not to use computational technology during class. On some days, you will be required to bring a personal computing device (e.g., a laptop computer), according to the schedule below. On remaining days, you will be given the choice as to whether to use computational technology as part of your learning process.

This policy is subject to revision, depending on student engagement over the course of the semester.

Academic Honesty

"If I have seen further, it is by standing on the shoulders of giants" (Isaac Newton, 1676).

In this class, you are both encouraged and will need to draw on the work and ideas of others. However, you must do so with appropriate acknowledgement. For scholarly writing, news media, books, or other publications, this usually means citation. In other cases, a footnote and/or an acknowledgement section may be more appropriate (for example, see acknowledgements made in some of the instructional material during class).

Plagiarism will ***not*** be tolerated. If in doubt, ask the Professor or see Lehigh's plagiarism policies (available from http://library.lehigh.edu/content/plagiarism_policies). Consider, also, the following Lehigh Student Senate Statement on Academic Integrity.

"We, the Lehigh University Student Senate, as the standing representative body of all

undergraduates, reaffirm the duty and obligation of students to meet and uphold the highest principles and values of personal, moral and ethical conduct. As partners in our educational community, both students and faculty share the responsibility for promoting and helping to ensure an environment of academic integrity. As such, each student is expected to complete all academic course work in accordance to the standards set forth by the faculty and in compliance with the University's Code of Conduct."

ACCOMMODATIONS

Lehigh University is committed to maintaining an equitable and inclusive community and welcomes students with disabilities into all of the University's educational programs. In order to receive consideration for reasonable accommodations, a student with a disability must contact Disability Support Services (DSS), provide documentation, and participate in an interactive review process. If the documentation supports a request for reasonable accommodations, DSS will provide students with a Letter of Accommodations. Students who are approved for accommodations at Lehigh should share this letter and discuss their accommodations and learning needs with instructors *as early in the semester as possible*. For more information or to request services, please contact Disability Support Services in person in Williams Hall, Suite 301, via phone at 610-758-4152, via email at indss@lehigh.edu, or online at https://studentaffairs.lehigh.edu/disabilities.

DISSENT

In this class, you will be asked to critique others' work, including both previously published papers and that of fellow classmates. Disagreements that arise in this process are both allowed and encouraged. However, disagreement must be voiced in a civil manner. Moreover, you should provide *constructive* critique. Saying something blatantly negative about another's work – that you do not like it, that it is bad, that it makes no sense, etc. – does not help that person. What can help is pointing out unfortunate, undesirable, or unintended consequences, as well as offering alternatives. Discussions of the readings will provide students with opportunities to practice such constructive critique before they will be asked to critique one another.

From the Lehigh Principles of our Equitable Community:

> *We recognize each person's right to think and speak as dictated by personal belief and to respectfully disagree with or counter another's point of view.*

Lehigh University endorses The Principles of Our Equitable Community. We expect each member of this class to acknowledge and practice these Principles. Respect for each other and for differing viewpoints is a vital component of the learning environment inside and outside the classroom. See http://www.lehigh.edu/~inprv/initiatives/PrinciplesEquity_Sheet_v2_032212.pdf

LEHIGH UNIVERSITY POLICY ON HARASSMENT AND NON-DISCRIMINATION

Lehigh University upholds The Principles of Our Equitable Community and is committed to

providing an educational, working, co-curricular, social, and living environment for all students, staff, faculty, trustees, contract workers, and visitors that is free from harassment and discrimination on the basis of age, color, disability, gender identity or expression, genetic information, marital or familial status, national or ethnic origin, race, religion, sex, sexual orientation, or veteran status. Such harassment or discrimination is unacceptable behavior and will not be tolerated. The University strongly encourages (and, depending upon the circumstances, may require) students, faculty, staff or visitors who experience or witness harassment or discrimination, or have information about harassment or discrimination in University programs or activities, to immediately report such conduct.

If you have questions about Lehigh's Policy on Harassment and Non-Discrimination or need to report harassment or discrimination, contact the Equal Opportunity Compliance Coordinator (Alumni Memorial Building / 610.758.3535 / eocc@lehigh.edu).

SCHEDULE

Readings listed on each day are to be done *before* that day. The astute reader will notice some days designated as [flex topic] days. These days serve two purposes on the schedule. First, they allow the instructor and the students to custom tailor the class to specific areas of interest that emerge over the course of the semester. Second, they ensure that the core material is covered in the event that any class meeting days need to be cancelled (e.g., for weather).

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
WEEK 1		
T – 8/27	WELCOME Assign discussion leaders. Review semester plan.	Instructor Leads Discussion
R	GETTING DATA Read the Syllabus before class. David Mimno. 2015. Data Carpentry. http://www.mimno.org/articles/carpentry/	Bring Laptop
WEEK 2		

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
T – 9/3	HOW TO READ A RESEARCH PAPER Paul Edwards. How to Read a Book. https://pne.people.si.umich.edu/PDF/howtoread.pdf David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. Computational Social Science. Science 323, 5915: 721–723. https://doi.org/10.1126/science.1167742	Instructor Leads Discussion
R	INTRO TO JUPYTER NOTEBOOKS Anne Bonner. 2019. Getting Started With Google Colab. *Toward Data Science*. https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c	Bring Laptop
WEEK 3		
T – 9/10	SOCIAL NETWORK ANALYSIS I Mark S. Granovetter. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78, 6: 1360–1380. Eric Gilbert and Karrie Karahalios. 2009. Predicting Tie Strength with Social Media. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 211–220. https://doi.org/10.1145/1518701.1518736	Instructor Leads Discussion
R	NETWORKX https://networkx.github.io/documentation/latest/tutorial.html	Bring Laptop
WEEK 4		

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
T – 9/17	SENTIMENT ANALYSIS James A. Russell. 1980. A circumplex model of affect. Journal of Personality and Social Psychology 39, 6: 1161–1178. https://doi.org/10.1037/h0077714 [*only read pages **1161-1167**, stop at the heading Multidimensional Scaling of Terms*] C.J. Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM)*, 216–225. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8109/8122	Student Leads Discussion
R	VADER IN PRACTICE Make sure to have VADER available in your Notebook environment: https://github.com/cjhutto/vaderSentiment Also see the following tutorials: http://www.nltk.org/howto/sentiment.html (skip to the section on VADER) https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f	Bring Laptop
WEEK 5		
T – 9/24	SOCIOLINGUISTICS I Kate G. Niederhoffer and James W. Pennebaker. 2002. Linguistic Style Matching in Social Interaction. Journal of Language and Social Psychology 21, 4: 337–360. https://doi.org/10.1177/026192702237953 Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of Power: Language Effects and Power Differences in Social Interaction. In Proceedings of the ACM Conference on World Wide Web (WWW), 699–708. https://doi.org/10.1145/2187836.2187931	Student Leads Discussion
R	CONVOKIT *Before Class*: Work through and be prepared to discuss one of the Notebooks linked from https://github.com/CornellNLP/Cornell-Conversational-Analysis-Toolkit/	Bring Laptop

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
WEEK 6		
T – 10/1	TOPIC MODELING I David M. Blei. 2012. Probabilistic Topic Models. *Commun. ACM* 55, 4: 77–84. https://doi.org/10.1145/2133806.2133826 Ted Underwood. 2014. Theorizing research practices we forgot to theorize twenty years ago. *Representations* 127, 1: 64–72. http://rep.ucpress.edu/content/127/1/64.abstract	Student Leads Discussion
R	INTRO TO TOPIC MODELING Install MALLET (http://mallet.cs.umass.edu/), r-mallet (https://cran.r-project.org/web/packages/mallet/index.html), and gensim (https://radimrehurek.com/gensim/) in your local environment.	Bring Laptop
WEEK 7		
T – 10/8	FRAMING Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. Journal of Communication 43, 4: 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x Eric P. S. Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri K. Gay. 2015. Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News. In *Proceedings of the Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 1472–1482. https://www.aclweb.org/anthology/papers/N/N15/N15-1171/	Student Leads Discussion
R	TEXT CLASSIFICATION Ensure that you have access to nltk (https://www.nltk.org/) and sklearn (https://scikit-learn.org/) in your notebooks environment.	Bring Laptop
WEEK 8		
T – 10/15	No Class Meeting – Pacing Break	

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
R	PROJECT PITCHES In class presentations of pitches	Project Pitch Due
WEEK 9		
T – 10/22	SOCIAL NETWORK ANALYSIS II: DIFFUSION Nicholas A. Christakis and James H. Fowler. 2007. The Spread of Obesity in a Large Social Network over 32 Years. New England Journal of Medicine 357, 4: 370–379. https://doi.org/10.1056/NEJMsa066082 Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural diversity in social contagion. Proceedings of the National Academy of Sciences (PNAS) 109, 16: 5962–5966. https://doi.org/10.1073/pnas.1116502109	Student Leads Discussion
R	SOCIOLINGUISTICS II: POLITENESS AND TOXICITY Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 250–259. https://www.aclweb.org/anthology/P13-1025/ Erin R. Hoffman, David W. McDonald, and Mark Zachry. 2017. Evaluating a Computational Approach to Labeling Politeness: Challenges for the Application of Machine Classification to Social Computing Data. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW: 52:1–52:14. https://doi.org/10.1145/3134687	Student Leads Discussion
WEEK 10		
T – 10/29	[Prof. Baumer away]	
R	[Prof. Baumer away]	
WEEK 11		

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
T – 11/5	FRAMING II William A. Gamson and Andre Modigliani. 1989. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *The American Journal of Sociology* 95, 1: 1–37. http://www.jstor.org/stable/2780405 Nona Naderi and Graeme Hirst. 2017. Classifying Frames at the Sentence Level in News Articles. In *Recent Advances in Natural Language Processing Meet Deep Learning (RANLP)*, 536–542. https://doi.org/10.26615/978-954-452-049-6_070	Student Leads Discussion
R	TOPIC MODELING II: GROUNDED THEORY Barney G. Glaser. 1965. The Constant Comparative Method of Qualitative Analysis. Social Problems 12, 4: 436–445. https://doi.org/10.2307/798843 Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? Journal of the Association for Information Science and Technology (JASIST) 68, 6: 1397–1410. https://doi.org/10.1002/asi.23786	Student Leads Discussion
WEEK 12		
T – 11/12	[Prof. Baumer away]	
R	[Prof. Baumer away]	
WEEK 13		
T – 11/19	PROJECT UPDATES Mini-presentations in class	Project Update Due
R	[Prof. Baumer away]	
WEEK 14		

DATE	TOPIC & READINGS	ASSIGNMENTS ETC.
T – 11/26	FAIRNESS [flex day] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. Retrieved from https://arxiv.org/abs/1808.00023v2 [***read only section 2.2 Defining Algorithmic Fairness***] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 656:1–656:14. https://doi.org/10.1145/3173574.3174230	Student Leads Discussion
R	No Class – Thanksgiving Break	

WEEK 15

T – 12/3	PROJECT PRESENTATIONS	In-class Presentations
R	PROJECT PRESENTATIONS	In-class Presentations Final written report **Due** Online Thursday, December 12 at 11:59 p.m.