SOC 102a: Analyzing Digital Media: Content Analysis of Online Text

Welcome

This course introduces students to computational methods for revealing patterns of thought in social media and other digital text data sets. **No prior coding experience is required.**

Combining social science and humanities questions with technical tools, students will learn how to collect, clean, analyze, and interpret text data (such as posts on social media) using exploratory methods such as text visualization, natural language processing, sentiment analysis, topic modeling, and network analysis. While technical skills are taught, this course focuses on how non-computer scientists can use computational methods to reveal patterns, limits, and omissions in cultural data, and how those results can be interpreted within social science and humanities frameworks.

This course **fulfills the University's Quantitative Reasoning requirement.**

Learning Goals

Core skills. By the end of the course we want you to have learned how to:

- Understand what constitutes digital text data and how it is produced
- Use basic computational tools to scrape, explore, and analyze social media datasets
- Apply sentiment analysis, topic modeling, and network analysis to research questions
- Interpret computational results critically, including ambiguous or inconclusive findings
- Develop a small, exploratory research project grounded in cultural or humanities inquiry

Knowledge. We also would like you to have learned to:

- Understand how communication helps to maintain social norms and institutions;
- Compare public online among communities to discern prevailing values and assumptions;
- Test hypotheses about the role of social media in social and political events.

Social justice, attitudes, and behaviors. We also would like you to:

- Reflect on the power that large social media companies can exert within the communities that use them;
- Identify patterns of thought that reflect biases within online communities;

- Design tools to identify disinformation and/or artificial generation of social media trends ("astroturfing.")

Components of Course Work

Homework & Assignments
Most of the homework assigned in the course will be to refine a project begun in class or to use a tool taught in class for your own work. Help will be provided in class and during student hours at the end of the week. Nearly every class will begin with a discussion and review of the assigned work.

Participation and attendance
Because this course is cumulative, attendance is essential. You may miss one class during the summer without penalty; after that, you will lose the opportunity to turn in the homework assigned during a missed class and thus will lose those points for your final grade.

Assessing Acquisition of Skills in Quantitative Reasoning
Students will learn basic statistics and how to study texts in a way that asks them to use statistical analysis to reach conclusions. Students will acquire these skills through exercises that ask them to compare word frequencies as distributions, collect and compare sentiment scores, looking at averages, variations, and differences across groups.

Students will also use network measures to compare the relative centrality of different terms and ideas within a text corpus. They will critically examine the outputs of their models in order to better understand the quantitative reasoning behind such models. They will also compare the results of applying different methods, such as LDA and BERTopic, to the same dataset, with the aim of showing them what each model captures and misses.

Evaluation and Grading

In each session, students will be introduced to methods or tools that they will use in class to develop their skills with the instructors' guidance. Students will then continue experimenting with the methods or tools by completing assignments on their own. This work will be submitted the day before the following class and will be evaluated for a grade. Student commitment to work in class will be graded as "in-class participation;" work outside class will be graded as "assignments" and "final project." Students will be given six assignments and will be allowed to drop their lowest grade. (All students will complete a final project).

Class Element	***Grade Percentage***
In-class participation	*10%*
Assignments	*70%*
Final Project	*20%*

Grades for the class: I will use the following to assign grades based on the course averages

while reserving the right to assign higher grades than those specified below.

>97	A+
> 93%	A
90-92.9%	A-

87-89.9%	B+
83-86.9%	B
80-82.9%	B-

77-79.9%	C+
73-76.9%	C
70-72.9%	C-

67-69.9%	D+
63-66.9%	D
60-62.9%	D-
<60%	E (F)

Weekly Schedule

Week of June 1 — Foundations & Exploratory Tools

Session 1 (Tuesday)

Introduction, Collaboration, and Ethics *(Theory + Discussion)*

Activities:

- Free write, discussion, and peer introductions
- Student reflections on:
    - Backgrounds and goals for the course
    - Comfort with programming or technical tools
    - Experience with evidence-based argumentation
    - Relationship to social media

Lecture & Discussion:

- Case studies: *What is text analysis used for?*
    - Ted Underwood, *"7 Ways Humanists are Using Computers to Understand Text"*
- Overview of course tools and expectations
- Ethical considerations in social media research
- Biases and limitations of metrics

Session 2 (Thursday)

Introductory Analysis Without Coding

Lecture:

- Structured vs. unstructured data
- What counts as a corpus
- Exploratory vs. interpretive analysis

Hands-On Practice:

Text Analysis Hands on with Voyant: https://voyant-tools.org/docs/#!/guide/tutorial

Bring texts to analyze!

Explore: Document Terms, Contexts, Corpus Terms, or anything else including any of the experimental tools such as Veliza or Dreamscape.

Pick your favorite and be ready to share to the group: a possible research question that could be generated from this text visualization.

Optional data visualization tool for exploration:

- Tableau

After-Class Work:

- Begin thinking about final project ideas
- Short reflective writing on possible research interests and datasets for final project

Week of June 8 — Python Basics & NLP Foundations

Session 3 (Tuesday)

Python Basics for Text

- Variables, strings, comments
- Python as a calculator
- Using Google Colab
- Introduction to working with text as data

In-Class Lab Assignment: Students create a short Python notebook demonstrating:

- Basic string operations
- Calculating text length
- Simple numerical ratios
- A short humanities reflection in Markdown

Session 4 (Thursday)

Python for NLP: Tokens and Word Counts

Lecture:

- Vocabulary, tokenization, dictionaries
- Functions and basic NLP logic

In-Class Lab Assignment:

- Build a tokenizer function
- Build a word-count function
- Demonstrate functions on a short text
- Interpret word patterns from a humanities perspective

Week of June 15 — Data Collection & Project Proposals

Session 5 (Tuesday)

Understanding APIs and Social Media Data

- What APIs are and how they work
- Demonstration of data scraping workflows
    - Scraping Bluesky and Reddit data
- Review of Python and NLP concepts

Session 6 (Thursday)

Dataset Acquisition & Preliminary Analysis

In-Class Lab Assignment:

- Collect a preliminary dataset using provided notebooks
- Organize and inspect data in spreadsheets
- Conduct exploratory close reading and metric-based analysis

After lab Assignment: Project Proposal
Students submit:

- Dataset description
- Guiding research question
- Humanities field and relevance
- Hypothesis
- Ethical considerations
- Preliminary literature connections

Week of June 22 — Modeling and Mapping Meaning

Session 7 (Tuesday)

Sentiment Analysis

- Conceptual overview of sentiment analysis
- How sentiment models work
- Strengths and limitations
- Biases and interpretive challenges

In-Class Lab Assignment:

- Apply VADER sentiment analysis to dataset
- Compute average sentiment
- Identify extreme cases
- Create a basic visualization

Students write short analytical reflections connecting sentiment results to their research questions.

Session 8 (Thursday)

Topic Modeling

- Introduction to LDA and BERTopic
- How topics are generated and visualized, using pyLDAvis

In-Class Lab Assignment:

- Run topic models on datasets
- Compare multiple outputs
- Reflect on what topic models reveal, and fail to reveal, about cultural meaning

Week of June 29 — Networks & Synthesis

Session 9 (Tuesday)

Social Network Analysis

- Nodes, edges, and graph types
- Centrality measures
- Network visualization platforms

In-class Lab Assignment:

- Generate network graphs
- Interpret relationships and structures
- Write a short analytical discussion connecting networks to research questions

Session 10 (Thursday)

Review, Integration, and Project Workshop

- Review of all technical methods
- In-class work time on final projects
- Individual check-ins and Q&A
- Guidance on selecting methods and interpreting results

Students continue refining projects beyond this session.

Final Project

Students complete an exploratory final project applying computational methods learned throughout the course to a research question. Demonstrate at least three visualizations and three calculations. Not all methods will produce clear or stable results. Learning how to interpret ambiguity, failure, and silence in computational outputs is a core objective of the course.

Important Course Policies

Expectations for student contribution to coursework
Success in this two-credit intensive course is based on the expectation that students will contribute at least 9 hours of study time per week in preparation for class.

Code of Academic Conduct
You are expected to be familiar with, and to follow, the University's policies on academic integrity. You are expected to be honest in all of your academic work. Please consult **Brandeis University Rights and Responsibilities** for all policies and procedures related to academic integrity. Allegations of alleged academic dishonesty will be forwarded to Student Rights and Community Standards. Sanctions for academic dishonesty can include failing grades and/or suspension from the university. **Citation and research assistance** can be found on the **university library website**.

Use of generative AI tools (University required text)
You may use AI programs e.g. ChatGPT to help generate programming code when explicitly allowed. However, you should note that the material generated by these programs may be inaccurate, incomplete, or otherwise problematic.

You may not submit any work generated by an AI program as your own. If you include material generated by an AI program, it should be cited like any other reference material (with due consideration for the quality of the reference, which may be poor).

Accommodations
Brandeis seeks to create a learning environment that is welcoming and inclusive of all students, and I want to support you in your learning. If you think you may require disability accommodations, you will need to work with Student Accessibility Support (SAS). You can contact them at 781-736-3537, email them at access@brandeis.edu, or visit the Student Accessibility Support home page. You can find helpful student FAQs and other resources on the SAS website, including guidance on how to know whether you might be eligible for support from SAS.

If you already have an accommodation letter from SAS, please provide me with a copy as soon as you can so that I can ensure effective implementation of accommodations for this class.

Respectful environment
Brandeis University is committed to providing its students, faculty, and staff with an environment conducive to learning and working, where all people are treated with respect and dignity. Please refrain from any behavior toward members of our Brandeis community, including students, faculty, staff, and guests, that intimidates, threatens, harasses, or bullies.

Missed classes/assignments
Because learning in the class is cumulative, we will not be able to accept late work. If you have a truly major health or family emergency, please contact us as soon as possible so we can make sure you stay on track.

Laptop computer and cell phone use
During class time, students may use laptops and cell phones for course work only.

Learning Management System (LMS) usage
We will use Moodle to post readings and other course materials and to collect assignments.