SOSC 4300 / SOSC 5500: Computational Social Science

Fall 2020
Lecture Time: Tuesday 9:00 - 11:50AM
Tutorial Time: Friday 6:00PM - 6:50PM

Tentative, this version prepared on August 30, 2020

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Teaching Assistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZHANG, Han</td>
<td>Peng, Wenwei</td>
</tr>
<tr>
<td>Office</td>
<td>Room 2379, Academic Building, Lift 15</td>
</tr>
<tr>
<td>Email</td>
<td><a href="mailto:zhangh@ust.hk">zhangh@ust.hk</a></td>
</tr>
<tr>
<td>Office Hour</td>
<td>Tuesday 2:00PM to 3:00PM</td>
</tr>
<tr>
<td>Zoom for office hour</td>
<td>TBD</td>
</tr>
</tbody>
</table>

### Prerequisites

- Students are expected to be familiar with the materials covered in basic statistics (e.g., SOSC 2400 for UG students and SOSC 5090 for PG students). Students with statistics knowledge but do not meet prerequisite can seek instructor’s approval for enrollment.

- Students should also have basic literacy in at least one statistical programming language. We will use R in tutorials. You can also use other programming languages such as Python, Matlab, Julia, etc., as long as you can finish course assignment and projects with the codes.

### Goals

Upon finishing the course, students should be able to:

1. Describe the opportunities and challenges of social research in the age of big data
2. Evaluate research on social phenomena from different fields, including social sciences and computer science/data science.
3. Practice the essential techniques to analyze social big data
4. Propose research questions that are suited to be examined by computational methods with big data
5. (For PG students): write a research article that utilizes the techniques and methods of computational social sciences to address social science problems.

### Grading

Your score will be accessed based on the following five components (no mid-term and final exams):

<table>
<thead>
<tr>
<th>Attendance and participation in class activities</th>
<th>UG students</th>
<th>PG students</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding exercises and short written assignments</td>
<td>10%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Literature review</td>
<td>20% (independent)</td>
<td>20% (independent)</td>
<td></td>
</tr>
<tr>
<td>Report</td>
<td>15% (3-4 people)</td>
<td>20% (3-4 people)</td>
<td>TBD</td>
</tr>
<tr>
<td>Presentation</td>
<td>5% (3-4 people)</td>
<td>10% (1-2 people)</td>
<td>TBD</td>
</tr>
<tr>
<td>Final Paper/Project</td>
<td>10% (3-4 people)</td>
<td>10% ((3-4 people))</td>
<td>Dec 1</td>
</tr>
<tr>
<td>Presentation</td>
<td>40% (3-4 people)</td>
<td>40% (1-2 people)</td>
<td>Dec 15</td>
</tr>
<tr>
<td>Write-up</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Attendance and participation in class activities**

- Based on class attendance and involvement in lecture and tutorial (e.g., ask questions during lectures and come to office hours).

**Coding exercises and short written assignments**

- These exercises test your knowledge of analyzing data using statistical software. Each exercise is due in **two weeks** after the release of assignment.

**Literature review**

Select a substantive research topic and summarize how the literature has used computational methods and/or big data are being used to study this particular research area.

- Some examples of research areas:
  - Sociological topics: internal migration, international migration, social inequality, race and ethnicity relations, happiness,
  - Political science topics: government performance, government policy (and its effectiveness), election, social movements
  - Economics: measuring economic growth with big data
  - History: anything related to historical phenomena
  - Psychology: measuring personality with big data
  - Communication and information science: fake news

- It is better for you to select a research areas that are similar to your final research paper. Students can discuss with instructors and TA for possible topics or feasibility.

- UG students should form groups of 3 or 4 participants. PG students can work individually or work with another PG student.

- Your performances will be accessed in two ways

  - **Report**: each literature review report should contain at least **10 pages, 12 points, double space**.
  - **Presentation of literature review (20 minutes)**: each student/group needs to present their literature reviews in class. Use the follow items for templates:
    * What is the research area you have chosen, and why it’s important
    * How people studied it traditionally (e.g., what data they use, what methods they use)
    * What are the advantages of using computational social science?
    * Some good examples
    * What are shortcomings of using computational social science methods to study this phenomena?
Final paper/project

Each student/group needs to choose a research topic and write a research paper (a conventional assignment) using computational social science methods or digital data. This research article should follow the format of a standard research article, with the following components: introduction; review of past studies; research methods and data; results; conclusions. UG students can form groups up to 4 people. PG students can work individually or work with another PG student.

- Presentation (20 minutes):
- Final paper/project: 20 pages, 12 points, double space, including Tables, Figures and References.

Alternatively, you can choose a project by building a demo/website/app that are useful for communicating some fancy social science data to laymen. Image your audience as people with no academic background, and you need to persuade them your stories with visualization and explanations. Some ideas of cool demo/projects can be found here: https://projects.fivethirtyeight.com/. If you choose this option, please discuss your ideas with the instructor at early stage. Your performance will be evaluated based on the following criteria:

- Presentation (15 minutes). Show case your project in front of the class. Focusing on whether your topic is interesting enough to attract your classmates’ interest.
- Final project: submit the project as a public accessible website/app that allow others (especially outside of the class) to access your output.

Grading policies

- Late delivery of due items will be marked down 75% if received within 1 day of the due date, and 50% if received within 3 days of the due date; you will receive zero credit if the due item is not delivered within 3 days of the due date. Contact the instructor if there are rare unforeseen circumstances.
- If you want to dispute a grade, please submit your argument in writing along with your assignment. We will evaluate the merit of your argument as well as perform a full reassessment of your entire assignment. This means that your grade may end up lower than it was originally.
- Final papers are checked by anti-plagiarism software. Students should take steps to avoid plagiarism and copying. For confirmed cases of plagiarism, severe sanctions including but not limited to a failure grade may be imposed.

Course Outline (Tentative)

The course materials will be drawn from lecture slides and assigned readings. Readings are available at Canvas. You are required to read the readings before the start of each class (except the first class). Optional readings are for students who are interested to read more on the topic.

Week 1 (Sep 8) : Digital Traces and Ethics

Optional readings


**Week 2 (Sep 15) Prediction vs. Explanation**


**Week 3 (Sep 22) : Survey**


Optional readings:


**Week 4 (Sep 29) : Text (I)**


Optional readings:


**Week 5 (Oct 6) : Text (II); Dictionary and Supervised**

Week 6 (Oct 13) : Text (III); Unsupervised And Word Embeddings


Optional readings


Week 7 (Oct 20) : Images and Spatial Data


Week 8 (Oct 27) : Network (I); small world


Week 9 (Nov 3) : Network (II); social influence vs. homophily


Week 10 (Nov 10): Network (III): simulations and tipping points

Week 11 (Nov 17): Causal inference: online field experiment


Week 12 (Nov 24): Warp ups; Presentations

Week 13 (Dec 1): Presentation of final paper and projects

**Tutorial Schedule (Tentative)**

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Sep 8)</td>
<td>Reproducible Research; Git and Project Management</td>
</tr>
<tr>
<td>2 (Sep 15)</td>
<td>Reproducible Research; literacy programming</td>
</tr>
<tr>
<td>3 (Sep 22)</td>
<td>Survey: offline and online</td>
</tr>
<tr>
<td>4 (Sep 29)</td>
<td>Text data: basic operations</td>
</tr>
<tr>
<td>5 (Oct 6)</td>
<td>Text data: dictionary counts</td>
</tr>
<tr>
<td>6 (Oct 13)</td>
<td>Text data: supervised learning</td>
</tr>
<tr>
<td>7 (Oct 20)</td>
<td>Text data: topic models</td>
</tr>
<tr>
<td>8 (Oct 27)</td>
<td>Network: basic operations</td>
</tr>
<tr>
<td>9 (Nov 3)</td>
<td>Networks: visualization</td>
</tr>
<tr>
<td>10 (Nov 10)</td>
<td>Networks: simulations</td>
</tr>
<tr>
<td>11 (Nov 17)</td>
<td>Networks: statistic</td>
</tr>
<tr>
<td>12 (Nov 24)</td>
<td>Online experiments</td>
</tr>
<tr>
<td>13 (Dec 1)</td>
<td></td>
</tr>
</tbody>
</table>