

The Data Science of Electronic Commerce

CAS CS 591 – Fall 2019

1 Project Guidelines

What am I looking for in a strong project?

- First, I am hoping that you are able to investigate an interesting component either of an online platform or of an e-commerce website in your work. In the first part of the class, we have largely focused on experimental measurement and modeling papers that begin from a compelling hypothesis and investigate the question through datasets. While I think this is an excellent model for a class project, it is not the only appropriate one. Other appropriate directions could involve design and analysis of a market mechanism, or starting a preliminary investigation of an open question, such as a theoretical question, in a previously published e-commerce paper.
- Second, I am expecting that there will be a rigorous quantitative evaluation of your approach, regardless of your methodology. This could take the form of an in-depth data analysis, or an experimental evaluation of a new method, or quantitative modeling of a new mechanism. What we are shooting for is a rigorously designed evaluation section that both cleanly and elegantly builds the case for the aspect or approach you elected to study. I am not expecting you to conduct your research in isolation – you should plan on frequent discussions with either myself or Harshal or other mentor to make sure your project is on track.

2 Deliverables

- Project proposal: by October 31, I ask you to submit a two-page summary (minimum) of your proposed project activities. This includes an overview of the area you intend to study and the research direction you intend to pursue, along with a basic description of your methodology or approach. Include a working bibliography (mandatory) of pieces of related work that you intend to draw from when conducting your study.
- We will reserve either the last class of the semester or coordinate with other CS courses project courses to set the time for a poster session of the class projects. I will expect your team to produce a poster (and accompanying 5-10 minute commentary) highlighting your results. We will talk about guidelines for a good poster presentation as the date approaches.
- During finals week, a conference-style writeup of your results will be due. I don't care a great deal about the length of the writeup (anywhere from 6-10 pages in double-columned format may be appropriate); the quality of the writeup is what you will be graded. We will talk about guidelines for a good paper as the end of the semester approaches. I will also give you a LaTeX template which I will suggest (but not require) you to use for your paper later in the course.

3 Project Ideas

Before I enumerate a few project pointers (none of which are guaranteed to bear fruit), let me first start with a word of caution. As with any piece of research, if you are not intensely interested in getting to the bottom of a problem, then you are not likely to make much progress. Therefore, **do not** choose a topic just because it has been suggested to you — pick a topic you care about deeply. (If there is no such topic, that’s a good indication that a career in data analytics or e-commerce research is probably not for you). Also, use the resources that are available to you, i.e. talk with me and with other faculty members about possible project directions that you are considering, and continue to consult with us throughout the semester. Finally, picking a good project can be very challenging, so get started thinking about topics and reading papers early. Some papers and datasets are listed below, but it is likely that the best projects will follow a line of the students’ own initiative, and not one of those mentioned below.

The Yelp Dataset Challenge Yelp has provided a ready-made project in its Dataset Challenge: <https://www.yelp.com/dataset/challenge>. Be sure to take the Yelp data in your own new direction.

Multi-Level Marketing: One interesting direction is multi-level marketing, where people at the top of the pyramid get compensated for recruiting others into (typically) a sales effort. A similar model was adopted by the winning MIT-based team that won DARPA’s Network (Red Balloon) Challenge. This paper is a really interesting read: Moshe Babaioff, Shahar Dobzinski, Sigal Oren and Aviv Zohar, “On Bitcoin and Red Balloons”.

Sometimes just seeing what data is out there might spark a new direction, but again, try not to reinvent the wheel. The Stanford Large Network Dataset Collection is one collection of datasets that might spark your curiosity (but some are getting older). A more recent dataset that I’ve used in my research is the 2017 Yellow Taxi Trip Data from NYC OpenData.

Online Diffusion: Many recent directions in electronic commerce related to social networks, and word-of-mouth marketing. One line is to reason about how information diffuses and propagates through these networks. Some strong papers include Flavio Chierichetti, Jon Kleinberg and Alessandro Panconesi, “How to Schedule a Cascade in an Arbitrary Graph”; and Eytan Bakshy and Dean Eckles, “Effects of Social Cues and Tie Strength in Social Advertising: Evidence from Field Experiments”.

Voter Models: A different social network direction relates to how individuals’ choices influence other individuals. A widely studied model is the so-called voter model, dating back to the early ’90s. Investigating this theoretical model in the context of consumer choice could make for a nice project bridging between theory and practice. Durrett et al. “Graph fission in an evolving voter model”. Proc. Nat’l. Acad. Sci. 109 (2012).

Online Advertising: Later in the class, we will be coming back to this topic in more depth, but there is a very rich ecosystem here. Just google “online advertising ecosystem kawaja” to see what I’m talking about. Many aspects of this field are open to investigation, for example: Mohammad Mahdian, Arpita Ghosh, Preston McAfee and Sergei Vassilvitskii, “To match or not to match: Economics of cookie matching in online advertising”.