Privacy in the Digital World: from Netflix to the Census
01:090:101:14

Time: Wednesdays (Sept 2 – Nov 4), 1:40pm – 3:00pm, 2020
Location: Synchronous remote via Zoom
Canvas: https://rutgers.instructure.com/courses/64491
Instructor: Ruobin Gong <ruobin.gong@rutgers.edu>
Office hours: by appointment

Seminar Description

The digital world we live in produces an explosive amount of personal data on a daily basis. From demographic surveys to biomedical studies, and to user information from massive online platforms such as Facebook and Netflix, large-scale collections of human data are translated into open-source databases or formats accessible for scientific research, as they play a crucial part in informing our collective biological and social functioning. While data transparency is key to fostering public knowledge and improving research reproducibility, it poses the risk of exposing the confidential information of respondents, patients and clients, violating their privacy. What measures can we take as a society to ensure effective learning from data, while at the same time respecting individuals' rightful claim to privacy?

In this class, we trace the historical and modern developments of privacy, and examine challenges around the execution of this concept from a multi-disciplinary perspective that integrates statistics, computer science, and economics. In a sense, privacy is the antithesis to information, thus the quest for privacy poses a fundamental dilemma to the growth of public knowledge. We weigh the social and policy implications of the inherent tradeoff between privacy and utility. We also take a close look at the recent development of a new mathematical framework for privacy, called differential privacy, and its adoption to protect the release of the 2020 U.S. Decennial Census. Philosophical implications of privacy on the broader subject of data ethics will also be discussed.

The seminar has no prerequisites, however prior exposure to probability, statistics or related topics is highly desirable.

Goals and Requirements

Attendance and participation (50%)
Your attendance and active participation is crucial to the success of the seminar. In order to have a fruitful class discussion, everyone must read the assigned weekly reading at least 24 hours before class meeting, and participate in the online discussion on Canvas by either asking a question, or answering a question from others. This is a *pass/no credit* course. More than two missed classes, or missed weekly readings, will result in a no credit grade. An "incomplete" grade will not be assigned in this course.

Leading a class discussion (25%)
From week 2 through week 10, a team of two or three students will form a "presenting group" and lead the class discussion. The goal of the presenting group is to organize the main ideas from the week's reading, and to present them in a way that facilitates a classroom discussion that evolve around the topics.

Scribing (25%)
Ingenious ideas sparkle through class discussions, and it would be wasteful not to write them down. Starting from week 2, the presenting group of the upcoming week will take scribe notes for the current week (i.e. week 4 group scribes for week 3, week 5 for week 4, ..., week 2 for week 10). The scribing group will share their scribe notes with the class after the meeting.

Absence Policy
Students are expected to attend all classes; if you expect to miss a class, please use the University absence reporting website https://sims.rutgers.edu/ssra/ to indicate the date and reason for your absence. An email is automatically sent to me.

Weekly Schedule of Topics and Reading

1. Introduction and course overview
2. Why statistical privacy?
   - Duncan, G. T., Elliot, M., & Salazar-González, J. J. (2011). Why Statistical Confidentiality?. In *Statistical Confidentiality* (pp. 1-26). Springer, New York, NY.
3. Breaching privacy
   - Narayanan, A., & Shmatikov, V. (2008, May). Robust de-anonymization of large sparse datasets. *In 2008 IEEE Symposium on Security and Privacy* (sp 2008) (pp. 111-125). IEEE.
4. Statistical privacy: traditional approaches
   - Abowd, J. M., & Schmutte, I. M. (2015). Economic analysis and statistical disclosure limitation. *Brookings Papers on Economic Activity*, *2015*(1), 221-293. [Section II only]
5. Privacy – accuracy tradeoff
   - Dinur, I., & Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 202-210). ACM.
6. Differential privacy
   - Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *In Theory of cryptography conference* (pp. 265-284). Springer, Berlin, Heidelberg.
7. Implications of differential privacy
   - Oberski, D. L., & Kreuter, F. (2020). Differential Privacy and Social Science: An Urgent Puzzle. *Harvard Data Science Review*, 2(1).
8. Transparent privacy
   - Gong, R. (2020). Transparent Privacy is Principled Privacy. *ArXiv:2006.08522*
9. Privacy and the 2020 U.S. Decennial Census
   - Sullivan, T. A. (2020). Coming To Our Census: How Social Statistics Underpin Our Democracy (And Republic). *Harvard Data Science Review*, 2(1).
10. Privacy and data ethics
    - Hand, D. J. (2018). Aspects of data ethics in a changing world: where are we now?. *Big data*, *6*(3), 176-190.

<u>Disability Services</u>
Rutgers University 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 the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: https://ods.rutgers.edu/students/documentation-guidelines. If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with a Letter of Accommodations. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. To begin this process, please complete the Registration form on the ODS web site at: https://ods.rutgers.edu/students/registration-form

<u>Academic Integrity</u>
Plagiarism is a major offense at Rutgers University. You are responsible for understanding the academic integrity policy and following these principles. Failure to uphold these principles of academic integrity threatens both the reputation of the University and the value of the degrees awarded to its students. Every member of the University community therefore bears a responsibility for ensuring that the highest standards of academic integrity are upheld. **The complete academic integrity policy can be found here: http://academicintegrity.rutgers.edu/.**