

1. Identification risk in anonymized data (4 points).

- a) I choose to describe health insurance records and credit card traction data in the table. Both datasets have re-identification attack problem. You cannot directly build connection to a specific person from the dataset, but it can still be potentially linked to a that person through the personally identifying information. For the health insurance records in Sweeney (2002), through the three specific common variables (zip code, birth date and gender), this data can be linked to the voter list data. Also, financial data sets contain a lot of personal information, such as shopping pattern and potential income. However, Montjoye et al. (2015) can still track the specific person by the time and place of credit card transaction data.
- b) In Sweeney (2002), the Group Insurance Commission (GIC) owns the health insurance data for state employees in Massachusetts. They can provide researchers with copies of these medical data, and even sell it to businesses because the records are anonymous. However, the information can still be linked by zip code, birth date and gender to a particular person in other data sets. In Montjoye et al. (2015), the author used 3-month credit card traction data of “1.1 million users in 10,000 shops (Montjoye et al. 2015: p.3)”. For each user, the author exacted the specific transactions in subsets and matched them based on other known information. Based on this approach, it is even possible to calculate the probability of a successful match.

2. Describing ethical thinking (3 points).

When this information leaks happened, Kauffman (Sep. 30, 2008b) said they “Were sociologists, not technologists”. For sociologists conducting social experiments, all they can do is collecting data as comprehensively as possible to avoid the experiment being affected by any uncertainties. This is exactly what those professors and research assistants did. According to the principle of Beneficence (Salganik, 2018: p.296), researchers should rely on cost-benefit analysis to maintain ethical balance. They did maximize their research result by collecting data on Facebook. However, although they are not technologist, they should still do a better job of protection before deciding to disclose the data.

Kauffman (Sep. 30, 2008b) clarified that all the data in their data set is from public sources on Facebook. This public information is available to anyone who has a Facebook account. According to Consequentialism (Salganik, 2018: p.301), informed consent is to help participants avoid some of the negative consequences of experiments that are not well designed. In this case, all the data and information contained are derived from publicly available materials that they voluntarily publish on social media. Therefore, even if some information is leaked, there will be no disastrous consequences for the participants. Under these circumstances, “a pure

consequentialist might be willing to waive the requirement for informed consent (Salganik, 2018: p.301)".

The explanation in Kauffman (Sep. 30, 2008c) can be regarded as the combination of the first two. Also, it pointed out the researches' ethical duty. According to Deontology (Salganik, 2018: p.301), request of informed consent is the embodiment of the researcher's respect for the independent will of the participants. In this case, the researchers did not ask for any other information about the participants without permission. Instead, if someone wants to work hard to crack this data, we may need to ask if he has ulterior motives.

3. Ethics of Encore (3 points).

- a) In Narayanan's and Zevenbergen's assessment of the Burnett and Feamster (2015) Encore study, there are still some debates about whether Encore is human-subjects research. In theory, the information collected on Encore is not individual information, but some data on the censorship. But on the other hand, whether it is human subject or not depends on whether it is possible to cause potential harm to the individual. Therefore, "the internet users may not be subjects per se, but they can still experience harm due to the research being conducted (Narayanan and Zevenbergen, 2015: p.13)".

Talking about the principle of Beneficence, Narayanan and Zevenbergen (2015: p.14) analyzed in detail the benefits and harms of Encore research and how to mitigate the potential harms. According to Salganik (2018: p.296), Beneficence is about the researchers could conduct the cost-benefit analysis and strike an appropriate ethical balance. The benefit of Encore is "enhancing the ability to create effective censorship circumvention tools." (Narayanan and Zevenbergen, 2015: p.15) The widespread censorship is not a bad thing in some ways. The assessment of the censorship should be more focused on "the transparency and accountability of decision-making regarding censorship". (Narayanan and Zevenbergen, 2015: p.16)

Besides, Narayanan and Zevenbergen (2015: p.14) talked about the possible risks of Encore. They mentioned several caveats that we should pay attention to. Firstly, "current online tracking practices" are seriously inconsistent with the expectations of the users. (Narayanan and Zevenbergen, 2015: p.18) Secondly, the scale and probability of harm depends on the type of censored website and the reason why the website is censored. Last but not least, there still exists "other types of harms that might result". (Narayanan and Zevenbergen, 2015: p.19) Then, they talked about several ways to mitigate the harm.

- b) I will access the ethical quality of the Burnett and Feamster (2015) Encore study

based on the four principles: Respect for Persons, Beneficence, Justice, and Respect for Law and Public Interest. Firstly, I think the researchers didn't show enough respect for the stakeholders. For some reasons, the individuals could be exposed to uncertain risk if their computers attempt to visit the blocked websites. Importantly, whether your web browser try to connect other websites is invisible to you. I think this seriously violates the individual's independent will. Secondly, the principle of Beneficence considers the safety of research participants and the balancing of potential problems. Due to the complex and flexible nature of internet, researchers are doing everything they can to minimize the risks that participants may face. "The current version of Encore tests only Twitter, Facebook, and YouTube, the rationale being that these domains are accessed regularly and automatically by most users' web browsers in the course of normal web browsing (Narayanan and Zevenbergen, 2015: p.19)." Thirdly, according to the law of Justice, burden and welfares should be distributed fairly. Narayanan and Zevenbergen (2015: p.16) even mentioned that participants should have corresponding benefits in the future while taking risks. Due to the uncontrollability of the internet, it is difficult for researchers to ensure that the benefits and risks are well distributed. Lastly, the Encore website contains the statement "Visitors of this page have performed XXX measurements of Web filtering" and also provides the options for the participates to choose not to take the future involvement. (Narayanan and Zevenbergen, 2015: p.21) Although this is legally required, they can still do more to strengthen this statement. For example, the reserachers must ensure that every participant has a good understanding of all statements. (Narayanan and Zevenbergen, 2015: p.21)

References:

Salganik, Matthew J., Bit by Bit: Social Research in the Digital Age, Princeton University Press, 2018, 294-301

Sweeney, Latanya, "K-Anonymity: A Model for Protecting Privacy," *International Journal on Uncertainty Fuziness and Knowledge-Based Systems*, 2002, 10 (5), 557-570.

Montjoye, Yves-Alexandre de, Laura Radaelli, Vivek Kumar Singh, and Alex Sandy Pentland, "Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata," *Science*, 2015, 347 (6221), 536-539.

Zimmer, Michael, "But the Data is Already Public: On the Ethics of Research in Facebook," *Ethics and Information Technology*, 2010, 12 (4), 313-325.

Kauffman, Jason, "I am the Principle Investigator...", Blog Comment, MichaelZimmer.org, <http://www.michaelzimmer.org/2008/09/30/on-the-anonymity-of-the-facebook-dataset/> , Sep. 30, 2008b.

— , "We did not consult...", Blog Comment, MichaelZimmer.org, <http://www.michaelzimmer.org/2008/09/30/on-the-anonymity-of-the-facebook->

[dataset/](#) , Sep. 30, 2008c.

Narayanan, Arvind and Bendert Zevenbergen, “No Encore for Encore? Ethical Questions for Web-based Censorship Measurement,” Technology Science, December 15 2015, 13-21