Homework #8

Angela Zorro Medina

1. Identification risk in anonymized data (4 points)

a. The papers I selected are Sweeney (2002) and Narayanan & Shmatikov (2008). In section "Understanding and managing informational risk" of Salganik's book "Bit by Bit," Salganik uses these two papers as examples of how anonymized data has a significant risk of re-identification. Salganik (2008) explains that the combination of different datasets or information sources allows identifying individuals' identities in anonymized datasets. Although, at first, the existence of those two datasets or information sources separately can be seeing as not risky, by merging them it is possible to re-identify the persons. This occurs when both information sources or datasets contain "key variables" that allow us to match observations in both datasets (one anonymized and the other one not-anonymized). These "key variables" work like a key that pairs observations when we have different "key variables" in both datasets the risk increases because it is harder to have multiple observations with the same values in different variables. In these two papers that was exactly the issue, there were "key variables" that give us unique identification of a person that it was possible to de-anonymized the datasets.

In both cases, Sweeney (2002) and Narayananan & Shmatikov (2008), the original datasets containing sensitive information were anonymized by the provider of the data (Group Insurance Commission and Netflix respectively). However, there is more information about those persons in the world that could help us match the two sources and identify each person. In the case of Sweeney (2002) the second source was another dataset set (voting records). In the case of and Narayananan & Shmatikov (2008), by applying an algorithm that matches different non-anonymized sources (e.g., IMDB reviews, colleagues information, etc.) with Netflix anonymized data.

b. In Sweeney (2002), the author got access to data from the Group Insurance Commission (GIC) containing anonymized data of thousands of states employees detailing their health records. Despite the efforts for anonymized the data[1] some of the variables in the dataset provided by GIC allowed the researcher to identify individuals using other datasets. In this particular case, Sweeney used voting records from Cambridge, Massachusetts to identify the information of GIC's dataset that belonged to the Governor of Massachusetts William Weld. Sweeney used the variables zip code, birth date, and sex, and was able to match the name, home address, party and date registered to vote with the anonymized medical records. The result, only one observation in the anonymized medical records had the same information of zip code, birth date and sex in the voting records, making possible to identify the medical records of William Weld. Having access to medical records could have important negative consequences in people's lives. Based on the literature that argues that some medical conditions caused stigma and discrimination in society, revealing medical records can put in risk of discrimination and stigmatization a vulnerable population (Greene-Shortridge 2007; Harangozo et al. 2013).

In the second paper, Narayanan & Shmatikov (2008), the anonymized data used was provided by Netflix. The authors created an algorithm that could use different types of sources to identify a person in the anonymized data. For example, Narayanan & Shmatikov (2008) used public records published in IMDB to

[1] GIC removed the names and the addresses of the employees.

de-anonymized Netflix data. Although we might be tempted to say that the information released by Netflix is not as sensitive as medical records, the lawsuit presented against Netflix in 2009, pointed out that some renting information can contain highly sensitive information. The lawsuit uses to examples, "Brokeback Mountain" and "The Passion of the Christ." According to the plaintiffs, by revealing that someone rented those movies Netflix could be giving personal details of a person's preference that she does not want to share because she can be stigmatized or discriminated (Doe v. Netflix).

Moreover, sometimes people watch documentaries or movies related to personal problems and medical problems, for example, obesity, mental health, pornography documentaries, etc., and by revealing that people watch those specific documentaries, Netflix is making public highly sensitive information. Additionally, when Salganik (2008) explains that I can use personal conversations to check the Netflix data of a particular person, everything becomes more problematic. Here, if companies are doing that check before hiring it could affect job opportunities or even cause someone to lose their job.

2. Describing ethical thinking (3 points)

Note: I understood this question as rewriting Kaufmann's comments using the principles and frameworks presented by Salganik. I did not understand this question asking me to respond to those claims, just framing them in another way without making any assessment. I answered this assuming that I was Kauffman.

First Comment

"While assessing the risk/benefit analysis, we concluded that there was no potential risk of identification, while the benefit of providing this data was significant for the sociological academic community. As Salganik (2008) points out the evaluation of potential risks requires technical, substantive expertise, that as sociologist we do not have. And because of this, we underestimated the potential risks of identification. Our estimation of the benefits come from our knowledge of the sociological discipline, and for a sociologist to know as much information of individuals as possible is an enormous contribution to the sciences. Using a consequentialists framework, by sub estimating the risk due to our lack of technical expertise we concluded that the benefits were greater than the risks, and we did not consider our obligation a "means" obligation but a "ends" obligation."[2]

Second Comment

"Although we recognize that we may have sub estimating the risk of re-identification of the data we published, what is the potential risk of identifying these persons? The data that we are using is already published on Facebook and is already available for those persons that want to see it and have the technical training to access it, like hackers. This means that we are not creating a new risk, the risk already exists and therefore our data does not have ethical problems in that sense. Moreover, the fact that we are not asking for consent to access the data that does mean that our research is ethically impermissible, in other words, the fact that

[2] Here Kauffmann fails to realize that he should have minimized the risks of potential re-identification in accordance to the beneficence principle. The researcher could have minimized the risk by changing the code in a way that it was not so easy to identify the college.

there might we some concerns regarding the Respect of Persons, that does not automatically make our study ethically problematic."[3]

Third Comment

"We comply with the principle of Respect of Persons by not using any other type of information besides publicly available information on Facebook. This means that people voluntarily agree to upload that information. Moreover, we did not make that information public, following the principle of Respect for the Law and Public Interest. We made our best effort to protect the identities of the persons in our datasets and did not incorporate any other private information to the public dataset, and we expected that the re-identification process was more complicated. In this case, we made our best efforts to minimize the risk by the principle of Beneficence".[4]

3. Ethics of Encore (3 points)

a. In 2015, Burnett & Feamster published a study in which they measured web censorship. Researchers introduced a system called Encore that "induced web clients to perform cross-origin requests that measure web filtering" (Burnett & Feamster 2015, p. 653), and allowed them to compile global-scale data of Web censorship. Based on those results, Burnett & Feamster (2015), analyzed the ethical concerns of their findings in the internet censorship discussion around the world. Narayanan & Zevengergen (2015) explain that every time a person visits a specific website, the Encore system executed a code without the consent of the visitors on their web browsers.

The lack of consent raised important ethical concerns about the Encore study. In fact, the ethical concerns regarding this study were so strong that when you access the paper, you find a statement from the publisher clarifying the ethical controversies that exist around Burnett & Feamster study. In their paper "No Encore for Encore? Ethical questions for web-based censorship measurement", Narayanan & Zevengergen (2015) discuss those ethical concerns and analyze them from different perspectives. First, Narayanan & Zevengergen (2015) analyze if this research project could be considered human subject research. Regarding this point, Narayanan & Zevengergen (2015) conclude that even though the center of the research was studying censorship systems and no individuals' behavior, then it cannot be properly defined as a human subject study. However, Narayanan & Zevengergen (2015) clarify that Burnett & Feamster (2015) should have followed the Menlo Report that advises "to respect individuals who are not targets of research yet impacted" and could have used alternative methods like using robots instead of human beings (Narayanan & Zevengergen 2015).

According to Narayanan & Zevengergen (2015), Burnett & Feamster failed into minimizing the potential harm that their study could create in the individuals involved. This does not mean that to Narayanan & Zevengergen

[3] Here Kauffmann fails to realized the potential harm that the identification can create and the ways they could have avoided it. By changing the code to make it less obvious would have been a great help. Moreover, the fact that other people is doing it or can do it does not mean that it is ethically acceptable. Following the principle of respect of law and public interest, researchers should try to comply with all the regulation even if other people is not doing it.

[4] Here Kauffmann fails to realized that the code they made public facilitates the identification of the data, which in fact increases the risk of people being identified with their data. When individuals consent to upload their information to Facebook they did not agree not make it public to other people outside that specific platform, which can be a violation of the ethical principles and laws.

(2015) what Burnett & Feamster did was completely wrong, in fact, they argue that it complies with all the rules and U.S. laws applicable to the specific case. The debate presented by Narayanan & Zevengergen (2015) exposes how complex are ethical issues nowadays with the use of new technological advances. While Burnett & Feamster (2015) did fulfill all the requirements that they had to comply, it is clear that they could have done some things better. For example, Narayanan & Zevengergen (2015) explain that by using robots not only Burnett and Feamster (2015) would have minimized the risk of harm individuals but also, they would have been able to increase the accuracy of their results by minimizing the biases related to measurement times.[5]

In light of the potential risk that exists and the lack of consideration of this risk by Burnett & Feamster (2015), Narayanan & Zevengergen (2015) analyze if the existing risk in the study was more than minimal risk. In this point, Narayanan & Zevengergen (2015) highlight the difficulties in identifying the potential risk in this case. Although Burnett & Feamster (2015) argue that they are not doing something different from what other agencies or individuals are doing on the internet and that the websites they were using do not create any potential risk because they were not particularly sensitive. However, Narayanan & Zevengergen (2015) explain that the problem with these arguments is that the definition of sensitive data can vary from the perspective of the researchers and the individuals that were used in the study. Additionally, even if other individuals are tracking people's information on the web that does not mean that "credentialed researchers and respected academic organizations" have the same ethical obligations that those other parties.

Finally, Narayanan & Zevengergen (2015) use a consequentialists framework to argue that Burnett & Feamster's (2015) study provide important benefits to society by bringing transparency to the web censorship conducted on the internet. In this point, Narayanan & Zevengergen (2015) discusses if it is worthy for the society to put those individuals in risk and if the benefits are fairly and equitably distributed. Narayanan & Zevengergen (2015) states that although under "western" point of view censorship is a negative thing for society this could not be true for other societies (for example, the Chinese society), and therefore the benefits from the results of the study cannot be distributed equally among the different societies that were studied. In conclusion, although under western standards unboxing censorship systems constitute an important benefit that would justify the risk involved in the study, in eastern societies (like Iran and China) this could not be true.

b. Although Narayanan & Zevengergen (2015) do a great job by pointing out the different ethical problems that exist in Burnett & Feamster's study, they limit their analysis to the consequentialist's framework. This means that the analysis done is limited to compare the benefit (the ends) of the study and the risks of harm to individuals. Narayanan & Zevengergen (2015) explain that the potential downsides of the study are that the authors did not minimize the risk of harm as much as they could have done it and that we do not have certainty regarding the equal distribution of benefits in Western and Eastern societies (when the study used both types of societies). However, I think that although a consequentialists approach is useful to identify potential problems and how to minimize the risks while maximizing the benefits, not always is possible to predict potential harms in the future.

[5] In other words, Burnett & Feamster (2015) could have done more to comply with the principle of beneficence in a better way.

In those cases, under which it is not clear that potential harm exists a deontological approach would be better to avoid future negative consequences for individuals. Using the deontological framework, Burnett & Feamster did not do the best they could to reduce risks. Under this framework there would not be a discussion regarding if it was valid not to required consent of the individuals in the study, there is only one answer: consent was required. Narayanan & Zevengergen (2015) exemplify how complex is the ethical dimension in this arena when using the consequentialist approach, and it is not clear who is receiving the benefits and what are the potential harms. This should be perceived as a case under which using the consequentialist approach lead to grey areas that created disagreement amount the academic community, and therefore in those cases, more straightforward criteria must be used. In this particular case, the deontological approach offers a unique obligation that the authors would have had to fulfill: consent. I agree that it could have created problems for the outcomes, but the outcomes of a research study should never be prioritized first than individual's protection. If social scientists are going to use human beings (as subjects or involved somehow in the study) we should follow strict requirements, because the dignity and safety of individuals should always be the priority. And if we believe that all persons are equally valuable and important then if there is a small possibility to harm at least one individual then we should desist from the research design and try another thing.

References

- Greene-Shortridge, T. M., Britt, T. W., & Castro, C. A. (2007). The stigma of mental health problems in the military. Military medicine, 172(2), 157-161.
- Harangozo, J., Reneses, B., Brohan, E., Sebes, J., Csukly, G., Lopez-Ibor, J. J., ... & Thornicroft, G. (2014). Stigma and discrimination against people with schizophrenia related to medical services. International Journal of Social Psychiatry, 60(4), 359-366.
- Jane Doe v. Netflix, (N.D. Cal. Mar. 19, 2010), Notice of Dismissal (Case No. C09-05903-JW-PVT)
- Salganik, M. J. (2018). Bit by bit: social research in the digital age. Princeton University Press.
- Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. Ethics and information technology, 12(4), 313-325.