Perspectives Assignment 8

Problem 1

Pick two of the examples in Table 1 and describe in one or two paragraphs how the re-identification attack in both cases has a similar structure.

a) I picked the studies on
   1) Demographic, administrative, and social data about students- Zimmer (2010)
   2) Phone Call MetaData- Mayer et. Al (2016)

   In both studies, we see instances of researchers or government agencies beginning with the correct intention of protecting individual users' privacy by removal of all 'Personally Identifiable Information' as defined by legal frameworks in the United States. This includes details such as the individual's name itself, affiliations with academic institution (such as student ID) or uniquely linked sequences (such as numbers in credit cards, drivers' licenses or social security registrations). However, either the accompanying information (such as the codebook for the publicly released dataset in Zimmer (2010) or accessible metadata (such as details of call duration, timing and parties involved in Mayer (2016)) can themselves provide sufficient clues for interested parties in identifying specific individuals. Their accuracy increases only further when this data is matched against other publicly available sources such as Facebook, Google Places, Yelp, etc.

   In both cases, interested individuals would need to submit an application for access to the this anonymized data through processes such as legal subpoenas to interested government agencies (for US phone metadata), or necessitating an agreement to the terms and conditions of use (for the Tastes, Ties and Time study in Zimmer (2010)). However, the 'metadata'- a codebook or phone logs for the two studies respectively- provided additional information to begin identifying individuals from the data set. Public data from websites and social media could be used either directly as the primary source (as for Tastes, Ties and Time) or for internal cross-verification (as shown by Mayer (2016) in their predictions on religion, home city, etc as compared to publicly stated information on Facebook).

b) In one or two paragraphs, describe how the data could reveal sensitive information about the people in the dataset for each of your two examples in part (a).

For Zimmer (2010), the sensitive information in question included sexual preference or political stances that had been included as part of broader sociological research. Due to data collection by of RAs from the same university as the participants in the study, the public

data set would include details from Facebook profiles that would have ordinarily been shown as private or unavailable for viewing by the general public. If matched with the specific individuals who furnished their sensitive data, there could be significant breaches of privacy and dangers of targeting.

This matching became possible elimination first at the level of identifying the academic institution, and then the individual students. The study description explicitly mentioned a North-Eastern American university, where the codebook documentation also furnished information on location (New England), funding sources (private), gender structure (co-educational), as well as the size of the first year cohort size (1640 students). This information alone narrowed the pool of possible colleges to 7. Among these options, only Harvard College offered the range of subject majors delineated in the codebook. The housing records included in the study, when compared against a publicly available video on Harvard's housing policies (the selection process of housemates by student preference in the freshman year, which may be correlated to their social graphs) only confirmed which college was being studied. Along with the 'cultural fingerprint' provided by cultural taste labels in the data, the limited representation of some nationalities and states in the cohort made their identification even easier.

In Mayer at al (2016), sensitive information included health conditions, purchase histories and religious affiliations. All three could be inferred by matching the numbers called (as per metadata) and the business category of those numbers (as listed publicly on Google Places and Yelp). In addition, the exact address of the phone user (which could be misuses for surveillance) and the categories of social relationships (for example, romantic partners) could be inferred from patterns and frequency of phone calls.

This study relied on a specially designed Android application that (with the user's own interest) would download both telephone metadata and facebook information. About 1 in 3 numbers- for both Individuals and businesses- could be easily linked back to their owners via use of publicly available APIs (such as Google Places, Yelp and Facebook). This accuracy could be sharpened easily by agencies with access to commercial databases. .

Problem 2

Upon the public announcement of this initial discovery, and general criticism of the research teams attempts to protect the privacy of the subjects, Jason Kaufman, the principle investigator of the T3 research project, was quick to react, noting that, perhaps in justification for the amount of details released in the dataset, `Were sociologists, not technologists, so a lot of this is new to us and `Sociologists generally want to know as much as possible about research subjects."' [Zimmer (2010) citing Kauffman (Sep. 30, 2008b)]

"From a consequentialist standpoint, we believed that this study would provide longstanding benefits to sociological research on the complexities of social networks. Though we live in a digital age, previous attempts at harnessing social media for these academic purposes have been limited to only the profile information listed on Facebook, small-scale surveys or ethnographies. The inherent self-reporting biases and limited depth of information have thus far hindered rigorous and comprehensive analysis. Our research overcomes these roadblocks. By following an entire cohort in a college across 4 years, and collecting as much data as possible, we believed that we would be able to discover hitherto inaccessible insights. Therefore, in the spirit of the principle of 'Beneficience' enshrined in the Belmont Report, we we would be maximizing the benefits gained from our research.

At the same time, we were- and still are- committed to the second aspect of this principle- of minimizing harm. We applied significant effort as per the highest standards of sociological research to minimizing any risks for the students in this study. The data has been anonymized with the removal of all personally identifiable information. Permissions and IRB review were formally introduced into our processes. Thus, all requirements of the Common Rule have been fulfilled. Further, the university whose students were involved itself granted us official permission to download their students' data. In this sense, we did not believe that informed consent from the students would be necessary.

Though we do not claim expertise as computer scientists, we have fulfilled all our obligations as social scientists. Any failings that may have occurred would only arise from our lack of understanding of the technology of online social networks, and certainly not a lack of precaution in terms of the design of the study itself or respecting ethical considerations or academic regulations.

No party is being harmed in the process of acquiring knowledge that would benefit society as a whole. Therefore, in our consequentialist view, the ends- of finally addressing longstanding sociological questions on networks through in detailed data collection on social media and university records- unequivocally justify the means."

[Kauffman] then attempts to diffuse some of the implicit privacy concerns with the following comment:

`What might hackers want to do with this information, assuming they could crack the data and `see' these peoples Facebook info? Couldnt they do this just as easily via Facebook itself? Our dataset contains almost no information that isn't on Facebook. (Privacy filters obviously aren't much of an obstacle to those who want to get around them")

"As researchers, our larger commitment is to ensure scientific rigour and reproducibility of our findings. Data sharing among other scientists can facilitate the expansion of knowledge and ensures accountability. Thus, sharing of our collected data is our duty, and offers large benefits to society as a whole. From a consequentalist standpoint again, we are justified in sharing this information. We also adhere to the principle of Respect for Law and Interest in the Belmont Report by ensuring transparency in how we arrive at our conclusions.

In the same vein, we would question what 'ends' would drive a hacker or any such individual hope to achieve by accessing this information, the majority of which is already available on Facebook? Any determined individual cold override some of Facebook's privacy filters. In such cases, such parties would be in violation of the Respect for Law, and are to be seen in this light.

We have not accessed any information not otherwise available on Facebook. We have not interviewed anyone, nor asked them for any information, nor made information about them public (unless, as you all point out, someone goes to the extreme effort of cracking our dataset, which we hope it will be hard to do)." [Kauffman (Sep. 30, 2008c)]

From a deontological perspective, we recognize the concerns that no matter what benefits our research provides, we are obliged to respect the autonomy of the participants in our study. Further, we concede that, like with any digital research, there exists information risk through sharing of data.

However, we have not shared anything that students haven't already shared publicly online. In a sense, they have implicitly 'consented' for this information to be available to the outside world. Here as well, we have employed every research-related safeguard available to ensure their anonymity- more than what would be provided on Facebook. 'Cracking' our dataset- by linking individual identities of students to their data- would prove forbiddingly difficult. As we just established, we do not see clear ends that would justify this level of effort.

Therefore, our beliefs on our research hold true primarily on consequentalist, but also on deontological grounds.

Problem 3

Assessment of Ethical Quality of the Burnett and Feamster (2015) Encore Study:

As mentioned above, the study focuses on a consequentialist approach (on gathering data on censorship which can prove nearly impossible through more conventional research studies), It almost completely sidelines deontological concerns on the means used. We may consider it in the light of the 4 principles enshrined in the Belmont and Menlo Reports (Salganik, 2018).

In terms of Respect for Persons, the study performs poorly by not taking any informed consent on the fact that research is being undertaken, let alone on the specifics of what is being measured. It does not fare well in terms of Justice too, since the potential gains are accessed by the exclusively by the researchers and not the individuals whose browsers carried the code entered. In terms of Respect for Law and Public Interest, the study does not fall under the purview of one unifying legal framework across the 170 countries where it was run. Though it may be have been accepted under extant US laws, a deeper inspection of cyber-security contraventions in other countries is essential. Most worryingly of all, the study fails on the ethical principle of Beneficience, since it subjects its participants to tremendous risks, since unassuming users' IP addresses may be intercepted by third parties involved in censorship, and may be incorrectly identified as engaged in espionage, sedition or in support of banned groups or movements. Thus, the Encore study sets out to achieve an important academic objective, but at too high a potential cost for its participants. The ethical justification for such approaches becomes particularly tenuous when considers other online research on censorship, such as by King et al (2013) which does not contravene the 4 principles outlined above to anywhere near the same extent.

Summary of Narayanan and Zevenbergen (2015):

Narayanan and Zevenberger are primarily concerned with analyzing the new paradigm of big data research on online censorship, and explore its implications specifically through the Burnett and Feamster (2015) 'Encore' study- where a snippet of code is invisibly introduced into a user's browser to access content from potentially filtered sites and then send back reports to researchers. They analyze this comprehensively on five grounds- the law, ethics, benefit-harm ratio and transparency. In their evaluation, the study achieves tremendous benefits minimal harm, while conceding that conceding that more safeguards could have been instated on the fronts of transparency and user consent.

The driving ideology here is a consequentialist one- that the ends justify the means (Salganik, 2018), provided that they lead to improvements in the state of the world. The ends here are to understand both the 'what' (content) and 'how' (technical mechanisms) of online censorship. After listing past research studies in the area, they conclude that in most cases, volunteers are needed within the geography where censorship is being studied, which would expose them to a degree of risk. This may have motivated the exploration of alternative 'means'- through a technical review enumerating technologies that have been used in gathering data and tracking online behaviour.

In its results section, it poses a number of critical questions. On exploring who would be considered the relevant stakeholders, it concludes the impossibility of the task given that users are distributed globally and the fundamental conflict between the distinct perception of 'scalability' in social science and computer science. Next, they explore whether these highly technical projects would be considered 'human subjects research', and concedes the challenge of conceptualizing the internet as a socio-technical system and thinking of the participants as individuals or merely technical devices

Another key principle at play here is that of Beneficience (Salganik, 2018) found in the Belmont Report and deeply rooted in the aforementioned consequentialist philosophy (ibid). As per this framework, researchers must conduct both a risk to benefit ratio analysis, and then ensure that this ratio may be considered ethically balanced. The authors concede that in such problematic cases, a deontological interest in informed consent for such (usually secrecy-imbued) network measurement research may not be feasible. They touch on two views- that censorship itself could be *universally* perceived as 'harmful', or where it may be beneficial in certain cultural contexts. The underlying implicit idea is that studies that improve society's understanding thereof could be considered beneficial.

In terms of the harm created by the research itself, the authors compare the Encore snippet to widespread use of malware and third-party trackers that an average user would struggle to monitor or control. They explore the differing levels of 'harm' possible based on the specific types of censored sites. It mentions the SIGCOMM committee's recommendation of ethical consent from users, over and above legal considerations (which Encore largely satisfies in the US.)

In summary, Narayanan and Zevenberger (2015) largely find the Encore study to be the first step in exploring nebulous ethics of online censorship research. From a purely consequentialist standpoint, their arguments suggest that risks are sufficiently minimal and the gains sufficient to uphold beneficience, and advocate for more explicit consent in related future research.

REFERENCES

King, Gary, Jennifer Pan, and Margaret E. Roberts. "How censorship in China allows government criticism but silences collective expression." American Political Science Review 107, no. 2 (2013): 326-343.