Yinxian Zhang

MACS 30000 Perspectives on Computational Analysis

Short Memo 1

Ethical Challenges for Social Scientists in the Digital Era

In 2008, a Harvard research team publically released its archive of an entire class of students at a northeastern university, including students’ Facebook profiles and their school official records of academic majors and residential dorms. This project soon came under scrutiny and has faced public condemnation. This paper analyzes the ethical issues of the Harvard project and its public release, particularly with reference to Salganik's four principles of ethical research. With the focal ethical controversy in mind, I then discuss how I think of the ethics of social sciences research, and whether I would use this dataset for my research if made available.

The most controversial parts of the Harvard project, in many people’s eyes, are related to its questionable data collection and protection process. First of all, the research team did not seek informed consent from their research subjects, referring to the fact that the data is already “public”, i.e. publically published on Facebook. However, as Zimmer (2010) points out, the Harvard team was hiring student research assistants to scrape data from Facebook, who may have privileged access to their fellow students’ Facebook profiles that are not necessarily open to the public. In other words, the data may not be as “public” as the research team claimed. In addition to problems associated with data collection, severer problems emerged when the research team made this data and its code book available to the public. Some scholars soon found that the alleged “anonymous data” is in fact identifiable based on information provided in the publically accessible code book, as least for a few students with unique characteristics.

Informed consent and anonymity (with regard to privacy protection) are thus the focuses of this ethical debate. They are also the focal points discussed by Salganik's (*forthcoming*) four ethical principles for research involving human subjects. The first principle, *respect for persons*, requires researchers to respect the autonomy of people, such that researchers need to obtain informed consent from participants. Apparently, the Harvard team fell short of informing research subjects of their project, let alone receiving consent. To be fair, it is debatable whether public data can be used without consent or awareness. Based on my own experience of applying for IRB approval, if the data is really public, meaning that anyone can access to the data without any form of permission(password, membership, etc.), then IRB is likely to waive the requirement of informed consent, provided that researchers can guarantee anonymity of participants. However, in the case of the Harvard project, the biggest mistake is to scrape Facebook data using some insiders’ (RAs) accounts – the data is no longer “public” if some students’ private setting only allows schoolmates/classmates to see their profiles. This move could be a rude invasion of privacy and a violation of people’s autonomy given such scenario.

The second principle, *beneficence*, requires researchers to weigh potential benefits and risks of the proposed study in order to reach a favorable and ethical balance. The Harvard project may not be directly beneficial to its participants, but other researchers and the social system may be greatly benefited from the knowledge on social networks accumulated by this project. Also the project per se is not particularly sensitive or harmful to its participants. However, the afterwards data sharing process has posed risks on its participants. Even though the researchers claimed that substantial efforts have been made to minimize the risks of identification, still some students have been identified after the data was released. The research team asserted that no students seem to have suffered any harm. One identified student seems to confirm this by saying that she doesn’t care (Parry, 2011). However, such identification risk bears great uncertainty – who knows what could be done after personal identities are leaked? Especially note that some Facebook profiles (including potentially sensitive information e.g. romantic and cultural tastes) and school records are not public! No harm so far does not mean no harm forever. The Harvard team failed to safeguard students’ information, posed risks on the participants, and thus violated the principle of beneficence.

The third principle, *justice*, addresses the protection of vulnerable people and stresses the distribution of the burdens and benefits of research. As discussed above, students in the Harvard research was not directly benefited, neither did they get financial compensation. By contrast, the scientific benefits go to academics and the society. This is certainly not an ideal distribution of burdens and benefits. Luckily that most participants in this research are not vulnerable in general sense. Taken together, the research team could have done better to honor the promise of justice, for instance, they can compensate the students by some means.

The last principle, *respect for law and public interest*, asks researchers to obey the law and hold themselves accountable by being transparent about their research. By and large, the Harvard team has adhered to this principle. First, they claimed that the data was downloaded with permission from Facebook and the given university, meaning that they considered and obeyed terms-of-services agreements of Facebook before scraping the data. Second, by publishing their work *Taste, Ties and Time* on academic journals and made their data available online, they enabled the broader community to gauge the ethical and academic issues of this project (and ironically led to controversy about privacy protection). If they provided better protection to the anonymous data, this project would have been a satisfying example of respecting for law and public interest.

Taken all together, this research provides us an opportunity to think about the evolving nature and expectations of ethics in online research. In the digital era, social scientists can obtain valuable data at large-scale, which was unimaginable in the past. But it is also associated with increasing ethical challenges. In part, I agree with Watts (2014) that this is a golden age of social science research, and we should not be intimidated by the controversy and anxiety emerging in the course of this social science revolution. Debate is *not* something we should avoid, but something good that enables us to reflect on and to improve our research. However, to respect and protect research participants should be the unchanging promise of scientific research. There are certainly bottom lines that we should never break.

In the case of the Harvard research, I can hardly imagine a sociologist resist the temptation of this perfect data, a data that could ease the long-standing difficulty of disentangling social selection and social influence mechanisms. I myself would be really excited if I can get access to the data. But would I be comfortable to use this data? Probably under certain circumstances: 1. inform the participants of my research and let them decide whether to withdraw; 2. only use information that was set to be “publically visible” during data collection. I won’t use the pieces of information that were only visible to the RAs. 3. Remove identifiers especially for people with unique characteristics. For instance, I won’t specify the home states of students, considering that there are only three students from Utah.

However, if I were a member of the Harvard team, I may probably not want to release the data again. I will make my research design as transparent as possible, but the original data will be kept away from the public, because it is almost impossible to anonymize social network data. After all, protecting participants is the priority for social scientists.

References:

Parry, M. (2011). Harvard Researchers Accused of Breaching Students' Privacy. *Chronicle of Higher Education.*

Salganik, Matthew J. (forthcoming) *Bit by Bit: Social Research in the Digital Age,* Princeton University Press, Open review edition.

Watts, D. J. (2014). Stop complaining about the Facebook study. It's a golden age for research. *The Guardian.*

Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology,* 12(4), 313-325.