Taste, Ties, and Time: An Ethical Review of the Harvard-Facebook Study

Chelsea Ernhofer

In 2008, a team of Harvard sociologists published a data set, created from the Facebook profiles of 1640 Harvard students (Lewis et al. 2008). The researchers, headed by Kevin Lewis and Jason Kaufman, employed research assistants to manually scrape data from the accounts of an entire incoming class, thus gathering a rich collection of information on social networking in a college environment. While the potential sociological discoveries are undoubtedly enticing, especially for social network analysts, doubts were raised concerning whether or not this study was entirely ethical. Going forward, I will address these ethical issues using Salganik’s four principles of ethical research (Salganik 2017). Ultimately, I will conclude that, because of ethical concerns, the data produced by Lewis and Kaufman is unusable.

The first of Salganik’s principles is respect for persons. This encompasses the idea that participants are autonomous beings and that any data collected should be done with their full awareness and consent. In regards to the Harvard study, however, this was not the case. Although the researchers gained the approval of Harvard’s administration and Facebook, the participants were never contacted in order to give consent. The authors of this research paper claimed that since the data was already publicly available on Facebook’s website, gaining consent was not a necessary step in order to gather their data (Zimmer 2010). This logic, however, is not entirely sound. As was previously stated, sociologists hired research assistants to comb through Harvard students’ Facebook accounts, pulling data from the profiles of the students in the desired cohort (Lewis et al. 2008). The assistants were instructed to only pull the data that they could view while logged into their respective accounts and to never contact participants for additional information. However, since the research assistants were members of the Harvard community themselves, it is entirely possible that they were able to access data that other individuals, outside of Harvard, could not see. This infers that perhaps the data was not public at all and only accessed through shared community ties. Moreover, the researchers were not using Facebook data alone but rather linking the online information found to housing records provided by Harvard University (Parry 2011). The claim that the data was already public and, therefore, up for grabs goes out the window.

In addition, researchers were confident that they would be able to completely anonymize the data, thus ensuring the privacy of participants. This hope was dashed when, days after the data was published, the ‘anonymous’ university being studied was identified as Harvard (Zimmer 2010). From there, it became possible to identify individual students, making the claim of privacy laughable.

The lack of privacy raises questions of the possible harm that this study inflicted upon its participants, especially when considering that information such as student housing records, cultural tastes, and approximated income was collected. This brings me to the subject of beneficence. Salganik describes beneficence as a researcher’s obligation to strike a balance between the benefits of a study and its potential harm (Salganik 2017). Studying social networks using Facebook ties certainly has benefits; the ability to collect data without measurement errors and use quantifiable measures of culture chief among them. However, when considering the potential harms to which the researchers subjected the participants and the lack of consent gained from the sample, I see no academic benefit that could sufficiently balance out the risks. This leads me to believe that the principle of beneficence was not met in the Harvard-Facebook study.

The third ethical principle, as outlined by Salganik (2017) is justice. This concept refers to both protecting vulnerable groups but also including minorities in research, the results of which could be beneficial to them. The sample for this research was an entire class of Harvard students – chosen, at least in some part, because of the percentage of freshmen students using Facebook (Lewis et al. 2008) – and as such, it does not appear as though researchers consciously sought out vulnerable groups but they did not make any attempts to protect vulnerable participants they found throughout the data gathering process either. Since their sample was pre-defined, it also became difficult to purposefully include minority groups in an attempt to gain sociological knowledge about them in particular. In general, it seems as though the researchers, while they did not openly challenge the principle of justice, were not concerned with observing it either.

The fourth and final principle is law and public interest. The researchers of the Harvard-Facebook study complied with all legal mandates in the gathering and publicizing of their data. As previously stated, they had permission from both Harvard University and Facebook to conduct their study. Lewis, Kaufman, and their team even had the support of Harvard’s Institutional Review Board (IRB), which, in theory, frees them from any legal backlash.

Transparency of research methods and intent also falls under the law and public interest umbrella and promises a slightly more complex discussion when applied to the Harvard-Facebook study than cut and dry regulations. Transparency refers to being clear about the goals of a particular research study and to take responsibility for one’s actions (Salganik 2017). I believe that the researchers of this study acted in a transparent manner and any information they kept private was in order to protect the identities of the participants. The very fact that this research was published in an academic journal supports the transparency and general honesty of its authors.

Rather than being purposefully unethical, I believe that the main problem with this research paper is the naiveté of the authors in regards to privacy. In my opinion, the authors had respect for the study’s participants and sincerely believed that they would be able to ensure their privacy, thus lowering considerably the potential harms that they could face. The data could have been made more anonymous by using more general fields of study and geographic locations, both steps which were later taken but not before it was too late.

For these reasons, I consider the Harvard-Facebook data to be unusable. The data set has already been linked to a very specific group, making it both unsafe for participants and unwise for researchers to use. Even though the data has now received a more thoroughly anonymizing treatment, the identities are already out there and further research would only put participants more at risk. I would, however, conduct research using new Facebook data if I could more confidently ensure the privacy and anonymity of participants through data encryption and more generalizable responses.

In summary, the data produced by the Harvard-Facebook study is unusable due to the major ethical missteps taken while gathering data. Researchers lacked respect for participants by choosing not to make them aware of their study, not seeking consent, but most importantly, not properly anonymizing the published data set. This led to an increase in risk on the part of the participants. Furthermore, the researchers did not show a sincere interest in justice, and while they did not break any laws in order to obtain their data, their naiveté in the privacy of said data cannot not be compensated for by obedience to law or transparency.

Works Cited

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks,* *30*(4), 330-342. doi:10.1016/j.socnet.2008.07.002

Parry, M. (2011, July 10). Harvard Researchers Accused of Breaching Students' Privacy. *The Chronicle of Higher Education*.

Salganik, Matthew J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Open review edition.

Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. *Ethics and Information Technology,* *12*(4), 313-325. doi:10.1007/s10676-010-9227-5