Book Review/ Compte rendu

Salganik, Matthew J. *Bit by Bit: Social Research in the Digital Age.* Princeton & Oxford: Princeton University Press, 2017, 448 pp, $35.00, paper, (9780691158648).

This book is addressed to two research communities: data scientists and social scientists. Its main goal is to help researchers from both camps "build their research intuition" in developing a research design in the digital age, when approaches to collecting and analyzing data are changing dramatically. Salganik argues that in this research era, researchers who deploy traditional social research methods face many obstacles, such as high proportions (up to 90%) of non-responses in phone surveys. The book also challenges probabilistic methodology, and argues that this is the right time for researchers to look towards adopting new methods. Salganik suggests that we must distinguish between custom-made data (those collected with well-established sociological methods purposefully) and readymade data (big data that are initially collected and stored not for social research but can be adjusted for the research purposes). Indeed, he does not advocate for replacing the traditional methods with new ones, but rather advocates for their complementary use.

There is an impressive volume of literature on big data and various research methods applicable to them, for example, *the Sage Handbook of Social Media Research Methods* (2017), edited by Luke Sloan and Anabel Quan-Haase. However, such books present collections of heterogeneous texts on the topic, while Salganik's book is a strict navigator in this field. Salganik understands big data broadly, defining them as any *digital traces* of human activity (13) recorded and stored through the help of digital technologies. Examples of this include information on purchases recorded by cashier machines in a store, information on electric power consumption held by the electricity companies, phone call records, taxi trip records, and so on. Salganik attunes us to the fact that oftentimes governments and businesses hold the big data that we as social researchers, can utilize.

Salganik discusses ten important characteristics of big data, some of them are generally positive for research, and some are negative. For instance, Salganik demonstrates that big data are non-representative. However, he argues that being non-representative does not translate into uselessness for valid scientific inference and prediction! Here, he refers

to John Snow's study of the 1853-54 cholera outbreak in London (29). At that time, it was believed that cholera spread due to "bad air," while Snow demonstrated on a small, non-representative segment of London neighborhoods that it spread through contaminated water. Another brilliant example of non-representative big data use is Wei Wang and colleagues' Xbox study (2015), that allowed us to predict the results of the 2012 U.S. Presidential election. The researchers used a non-probabilistic sample consisting of Internet gamers—the overwhelming majority of whom were young white men—and generalized the results to the whole U.S. population using advanced statistical methods (102-106). The author of *Bit by Bit* provides many more examples in order to destroy our stereotypes about sampling and representativeness, because in our minds, representativeness is tightly associated with scientific validity.

Each chapter of Salganik's book is dedicated to a particular research method applied to big data. Chapter 2 tells us about observing behavior, and about new horizons that are open to the researchers who can use big data for testing fundamental theories. Salganik consistently stresses the importance of creativity and on "building intuition," and the book is the collection of outstanding examples of studies that demonstrate, inspire and trigger that creativity. According to Salganik, the opportunities are endless, and researchers just need to train their brains to recognize them. For instance, Henry Farber (2015) tested two competing theories in labor economics (behavioral economic model and neoclassical model) having a dataset of working hours and hourly wages of New York taxi drivers, that was collected by electronic meters that the city requires taxis to use (42). Thus, again we see that big discoveries can be hidden in unusual places!

Chapter 3 discusses surveys and principles of asking questions in order to collect data. Here, Salganik presents a historical account of surveys. He takes the *Literary Digest* fiasco in predicting the 1936 U.S. Presidential election to analyze the main rationales of sampling, and demonstrates why after this grandiose failure, social scientists had many prejudices against non-probabilistic samples. He shows how the order of asking questions, the manner of asking them, and the presence of an interviewer can influence survey results. In the digital age, the human factor of influence can be mitigated. Using new technologies, studies with high predictive accuracy can be conducted. Very often, datasets obtained through different methods (traditional and "digital") are combined in one study to verify results and amplify their validity. This is the case with Joshua Blumenstock's study (2014), which examined poverty levels in Rwanda and combined phone call records obtained from the phone

companies (big data) and direct phone surveys of a small portion of cell phone users (122-130) to predict and measure household possessions.

Chapter 4 describes the experiment method, most often deployed in university laboratories where the most typical subjects are undergraduate students (weird subjects under weird conditions performing weird tasks), while field experiments are rather rare and costly. In the digital age, opportunities for field experiments are diverse and multiple. Salganik became convinced of the great advantages of the digital age methods as a graduate student, when more than hundred people participated in his own online experiment aimed at exploring the destiny of cultural products, while he was sleeping!

Chapter 5 is dedicated to a new method not yet fully adopted in social research—mass collaboration. Wikipedia is one such product of mass collaboration. The strongest advantage of this method is the opportunity to attract many collaborators, often for free, when the task is clear, but extremely voluminous, or when the researchers cannot find a solution to a problem, but believe that others may have a solution. If the project is presented to the public as financially attractive (i.e. Netflix Prize), or is joyful and entertaining when the project is designed like a game, it is possible to expect cooperation of many volunteers who will contribute their effort to the success of the study. The classical example provided by Salganik is the Galaxy Zoo project (2007), when volunteers were invited online to identify the color and shape of millions of photographed galaxies.

With the methods discussed above there is a huge concern about research ethics (Chapter 6). Due to digital technologies, it is now much easier to obtain diverse data than it ever was before, but it is harder to prove that it was done ethically. For example, in using mass collaboration how should the merits of the discovery be distributed among the many contributors? In the case of experiments, surveys and observation, how do we receive informed consent if our samples are comprised of millions of people? Thankfully, in this chapter, Salganik provides readers with convincing examples and clear instructions on how to solve these problems.

Chapter 7 summarizes the main ideas of the book, among which are the blending of readymade and custom-made strategies of research, participant-centered data collection, and the importance of ethics in digital research design.

Given the book's breadth, it is a recommended read for all scholars interested in the role that the internet and big data can play in social research. Every chapter is accompanied by a detailed mathematical commentary related to the analyzed method, recommendations for further

reading, and by the range of tasks that can be fulfilled in the class. Thus, Salganik's book can be successfully used as a manual for courses in computational social science, as well as for social science methodology classes.

University of Ottawa Tatsiana Amosava

Tatsiana Amosava's research focuses on nostalgia for the Soviet past in internet Russian-speaking online communities on three different websites with similar properties, including Facebook, and two Russia-based websites ("Odnoklassniki" and "VKontakte", which are politically idiosyncratic). Her work explores: what exactly are people nostalgic for, and why are they engaged in those activities online? Do specific groups and sites make a recognizable nostalgic subject? How does the quality of each site and forum shape the audience, and vice versa? To date, she has published two books on the sociology of language, and the official bilingualism in Belarus and Israel.