Big Data in Social Sciences
Matthew N. O. Sadiku, Mahamadou Tembely, and Sarhan M. Musa
Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX 77446, United States
Email: sadiku@ieee.org; mataembely@pvamu.edu; smmusa @pvamu.edu

Abstract—We are in the era of big data. Big data provides unprecedented opportunities to study human behaviour. Although large datasets alone are not sufficient for solving the most pressing of human problems, they can reveal patterns of human behaviour that are hard to observe in the laboratory. Social scientists have been lagging behind their peers in the physical and medical sciences in embracing big data. This paper briefly presents how social science research and teaching are evolving to meet the challenges and opportunities of big data.

Keywords—big data, social sciences, social media, big social data, computational social science

I. INTRODUCTION

Today’s world is experiencing a data revolution and it permeates almost every aspect of modern life. Data is being produced and stored at an unprecedented rate. The promise of the big data revolution is that in these data are the answers to fundamental questions of businesses, governments, natural sciences, and social sciences [1]. Businesses are learning that making effective, wise use of their data can impact their bottom line. Sociologists, historians, political scientists, economists, philosophers, and anthropologists are well placed to deal with the big data.

The increasing availability of big data has divided public opinion into two groups: the skeptics, who question the legitimate use of that data on the basis of ethical concerns, and the enthusiasts, who focus on the transformational impact of big data. Research that is exploiting big data is the most powerful proof of their transformational impact [2].

Although an enormous number of studies have been carried out on big data, its application in social sciences is recent. Big data has been increasingly used in social sciences to reveal individual differences and group dynamics. It has the potential to transform the way applied social science is done. Social scientists should understand the potential of big data in complementing traditional research methods.

II. CONCEPT OF BIG DATA

One may define big data as anything too big to fit into the analyst’s computer. Big data is when your method breaks down, when you need a completely new method to analyze the data. One may also define it as data with high volume, high velocity, and great variety.

The types of available data fall into various categories: social data (e.g., Twitter feeds, Facebook likes), data about geospatial locations (e.g., sensor data collected through mobile phones or satellite images), data collected from government administrative sources, only to name a few.

Figure 1 The 5V of big data [3].
The traditional characteristics of any big data include volume, variety, velocity, veracity, and value, as shown in Figure 1 [3]. Volume refers to the exponential growth of social data. Variety relates to various types (text, image, video) and forms of social data sources which may be structured, semi-structured or unstructured. Velocity refers to rapid speed at the social data is generated. Veracity refers to the truthfulness of data, i.e. whether the data comes from a reputable, trustworthy, authentic, and accountable source. Value of data depend on the characteristics and knowledge of the decision-makers, who are regarded as private or public actors who make decisions [4,5].

In addition to these traditional characteristics, “big data” also has the qualities of being relational, aggregated, multilevel, and merged. Relational data have linkages between individuals and groups; aggregated data include data that are combined from multiple sources; multilevel data combine measures at both individual and group levels [6].

III. SOCIAL SCIENCES AND BIG DATA

Social sciences include economics, public policy, sociology, management, and psychology. Typically, social scientists are interested in describing the activities of individuals and organizations in various economic and social contexts [7]. Social media and the underlying networks are becoming important in the social sciences. As the volume, variety, and complexity associated with the data can be difficult to manage, the study of social media data is challenging [8]. The challenge involves making sense of the exponentially increasing amounts of data being produced and stored for advancing new discoveries in social sciences.

Traditional social science techniques such as surveys (interviewing people or questionnaire) and laboratory experiments allow social scientists to carefully design their studies and determine how to measure variables of interest. The big data in the social sciences comes from social media, computer-mediated communication, administrative processes, financial transactions, and so on. Social networks, websites, smartphones, and various online applications constantly generate data on an unprecedented scale. Several research-relevant data sources are not available from the publicly but from data holders such as Facebook, Google, or governments.

The data itself is neutral and relative; it is what we do with it that makes it relevant for social science research [9]. Crunching more data enables us to take better decisions and solve social and political problems. Big data presents great opportunities to understand human behaviour on a large scale. The opportunities associated with moving big data to knowledge are relevant to the social sciences in two ways. First, social science communities are joining other scientific disciplines in revisiting the roles of theory, publication, data sharing, intellectual property rights, and knowledge accumulation. Second, social science research is needed to understand how big data assets can be created, accessed, shared, and utilized for advances in knowledge across disciplines [10]. Social scientists have been instrumental in developing the methods and techniques of large-scale data collections and analysis.

Big data researchers must carefully examine big data and remove the noise contained in the data. They also need to use proxies for their variables of interest. They can apply data analysis procedure (such regression, reliability tests, factor analysis, and multilevel analysis) to uncover the dynamics of social and psychological processes [11].

IV. BENEFITS AND CHALLENGES

The benefits of big data for social sciences and social impact are many. Big data has the potential to enhance the scientific method. Massive datasets and social networking sites provide opportunities to design experiments on a scale that was previously impossible in the social sciences. The heterogeneous and complex nature of big data allows researchers to include more relevant variables such as time, location, or population density than in traditional laboratory studies.

However, several challenges need to be addressed before ground-breaking discoveries can be made. The strategies for storing and analyzing big data in social sciences are complex and not without difficulties. At all stages from data acquisition to data interpretation, the social scientist must consider the ethical ramifications of their work and respect privacy, proprietary, and confidentiality. Current studies rely on one single data source such as Facebook or Twitter due to the difficulty of matching users from different platforms. As data becomes more and more valuable, ownership becomes important.

V. CONCLUSION

Big data is now a commodity in the social sciences and a cutting-edge topic. Due to big data, the research content of social science is going through a drastic change. The analysis of big data in social sciences opens up new avenues of research and makes it possible to answer questions that were previously unanswerable. Remarkable progress has been made in social and psychological sciences through the analysis of big data. There is need to train students in the social sciences to help them take full advantage of the exciting new opportunities big data brings.
Big data by itself is not a panacea for all challenges modern societies face. It will not transform the social sciences, but we have a lot to gain from the technologies that it offers [12]. Engineers, computer scientists, and social scientists have all converged on big data, creating opportunities for collaboration. As a result of this convergence, we are witnessing a potential shift from methodological individualism to methodological transactionalism and an openness to a wider range of social theories concerning interaction, attention focus, and meaning usage [13]. More information on big data in social sciences can be found in [7] and the two journals devoted to big data: *Journal of Big Data* and *Big Data & Society*.

REFERENCES


ABOUT THE AUTHORS

Matthew N.O. Sadiku (sadiku@ieee.org) is a professor at Prairie View A&M University, Texas. He is the author of several books and papers. He is a fellow of IEEE.

Mahamadou Tembely (mtembely@student.pvamu.edu) received his doctoral degree from Prairie View A&M University, Texas. He received the 2014 Outstanding MS Graduated Student award for the department of electrical and computer engineering. He is the author of several papers.

Sarhan M. Musa (smmusa@pvamu.edu) is an associate professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.