NEUE SCHRIFTEN ZUR ONLINE-FORSCHUNG

Cathleen M. Stuetzer / Martin Welker / Marc Egger (Eds.)

Computational Social Science in the Age of Big Data

Concepts, Methodologies, Tools, and Applications

HERBERT VON HALEM VERLAG

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists the publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Cathleen M. Stuetzer / Martin Welker / Marc Egger (Eds.)
Computational Social Science in the Age of Big Data.
Concepts, Methodologies, Tools, and Applications
Neue Schriften zur Online-Forschung, Band 15
Köln: Halem, 2018

ISSN 1865-2638

ISBN (Print): 978-3-86962-267-5
ISBN (PDF): 978-3-86962-268-2

http://www.halem-verlag.de
info@halem-verlag.de

TYPESETTING: Herbert von Halem Verlag
EDITOR: Imke Hirschmann, Köln
PRINT: docupoint GmbH, Magdeburg
COVERDESIGN: Claudia Ott, Grafischer Entwurf

Inhalt

CATHLEEN M. STUETZER / MARTIN WELKER / MARC EGGER

Big Data Analytics: Obstacles and Opportunities for Social Science

Digitalization has already permeated all areas of life and is omnipresent – not only in our everyday life, but also in politics, business and especially in science. Industry is now talking about wearables, the Internet of Things and industry 4.0 when it comes to promoting digitalization and digital networking within the production cycles. Over the past years, the velocity of technolgical advancements has tremendously increased that the exploitation of the so-called *digital footprints* – which we leave permanently in our everyday life – becomes the subject of social science research.

In the last 15 years with the emergence of social media platforms, we observe a new phase of data revolution (LAZER et al. 2009; ALVAREZ 2016). With the massive increase in data production Big Data is more and more discussed as *socio-technological* phenomenon (BOYD/CRAWFORD 2012; LAZER et al. 2009; ALVAREZ 2016). With the help of computational techniques the collection and extraction of data seem often easy to obtain and cheap (KING 2016). Thus, in social science we notice a computational turn in research, and that is mostly associated with high expectations on scientists (BOYD/CRAWFORD 2012). But how can we benefit from analyzing big (social) data? How can we handle the new data? Which analytical approaches, techniques, and instruments are actually used and discussed? Which skills are needed in that upcoming field? What should we know about ethical and privacy issues? And what are the consequences for theoretical considerations?

Although research on Big Data has a long tradition, only with the emergence of online communities and social media platforms in 1990s' Big Data

arise as socio-technological phenomenon (BOYD/ELLISON 2007). Actually, Big Data is a term which influences all fields of (applied) research, and, according to Boyd and Crawford (2012), »Big Data not only refers to very large data sets and the tool and procedures used to manipulate and analyze them, but also to a computational turn in thought and research« (p. 665). But what is the meaning of this computational turn in thought and research?

Traditionally, researchers are focused on answering research questions regarding a special target group. In the context of Computational Social Science (CSS), data is often captured first. Thus, the process of exploration of massive social data plays an important role to identify a focus group and generate hypothesis as well as research questions after that. It seems to be a fruitful opportunity to get new insights of social phenomena.

New perspectives on data analytics bring a lot of obstacles and opportunities for researchers. First of all, a new paradigm raises high expectations. According to Alvarez (2016) we »examine the social world in new ways« (p. 4) and bring Big Data to life. Second, we have always sought answers to the question on how the (social) world is constructed, and we tend to reflect new approaches critically. Third, with the emergence of social technology we need suitable infrastructures – not only technological and methodological but also socio-cultural. The opportunities are obvious. CSS based on theoretical approaches which explain social world as a connected world in different stages. From this point of view, we get insights about social processes related to communication, interaction, and social relations on the (social) web at different levels. By analyzing a new type of data and by using new techniques, we can extract new types of knowledge (LAZER et al. 2009). Continuing to this, we identify Big Data as a »social construct« to handle (social) phenomena within the connected world.

CSS as methodological approach offers systematized ways by using new techniques for collecting, extracting, and analyzing large-scale (social) data in (applied) research. CSS enables to track digital footprints and stream »social« traces with computational methods (LAZER et al. 2009; ALVAREZ 2016). New analytical approaches behind CSS allow to explore social behavior and the dynamics of causal correlations. CSS highlights opportunities in representing projections of the social world by analyzing digital traces people left behind (CIOFFI-REVILLA 2014). Most social traces are left using the Internet – but social science methods focused on human traces are older than the web (e.g. JAHODA/LAZARSFELD/ZEISEL 2015 [1933]). Nevertheless, nowadays we have a set of new possibilities reaching far beyond

the methods and instruments already applied in the 1930s. Technologically driven changing communication habits change the circumstances under which social research can sensibly be conducted in contemporary society. Therefore, social scientists ask at least three intrepid questions when it comes to the field of digital methods and algorithms:

1. Does the emerging research field of Computational Social Science require a new methodology?
2. Does it need new, conceivably additional, scientific quality criteria? So, does Computational Social Science require new or modified methods within current methodology or is there a demand just for better performance in practice?
3. And if we need new or modified methods, what could it be, which new or improved computational methods are prerequisite to the field? How is an optimal integration of computational methods achievable?

The classical research procedure now shifts towards data collection, data processing, and data storage. Extremely, the course is turned upside down: assumptions and hypotheses are discussed at last, after outcomes are generated with theories totally eliminated. A new methodology could reflect this approach at a meta-level critically. Secondly, this new empirical field is strongly based on computer science and its subdisciplines. But humanities and social science have another focus while computer science is often focused on the optimization of machine processes. The discussion of algorithms for research due to methodological and ethical problems mirrors these different rationalities.

Under those circumstances, does Computational Social Science expect an original set of theories or a theoretical framework? In other words, does CSS require a theoretical basis that links all the diverse works and studies and serves as a common draw of the field? And if yes, what framework could we use? A theoretical framework would be desirable but perhaps not necessary. As Schroeder and Taylor (2015) and Stegbauer (2009) have shown, Big Data studies on Wikipedia are extremely heterogeneous. The Big Data studies on Wikipedia had no common perspective but a common goal: to be able to answer research and professional questions by applying Big Data. In this context, for example, innovation theories possibly are most suitable, also approaches of common understanding, while theories of action may be less appropriate. But how do perform the new field in practice? Does it meet the accuracy and precision scientists and society need?

Accordingly, large multinational telecommunication companies and online service providers install their own departments of data, thus making science proprietary instead of publicly debatable and verifiable. The digital market leaders like Apple, Google, Amazon, Facebook, Microsoft, etc. set up their own research basis, storing data and using their platforms as massive research tools. But what is the impact of this trend on academic social science? What are the consequences?

Nevertheless, the potentials of introducing CSS as research paradigm in social science become visible. As the learnings of Online-Research and of 20 years of experience in validating new methods and instruments show us, it seems that CSS comes with a certain momentum to justify a new theoretical and methodological basis. The use of new ways of (social) data collection, extraction and exploration open up completely new fields of activity. We are at the beginning of a new era in applying new research theories, methods, and applications in (applied) social science.

With this book we want to initiate a discourse – theoretical as well as analytically – about the upcoming field of research. We present selected contributions to demonstrate the computational turn in social research as well as the relevance for the applied market research. Our selected contributions in this book underpin that CSS is a young and highly interdisciplinary field of research that primarily aims to generate complex data to usable information. The exploration of massive data is not only interesting for computer scientists, physicians, meteorologists, and/or business economist but also for psychologists, social scientists, communication experts, and political scientists. On the one hand, it is attempted to gain insights into social phenomena from process-generated data – on the other hand, new methodological approaches are used to answering questions about (social) impact mechanisms. This field of research is driven by the primary research mission to contribute further developments of evidence-based behavioral and impact research.

Within the first chapter »Epistemological Perspectives«, the authors open the discussion about the relevance for CSS as research field. They highlight obstacles and opportunities of Big Data analytics in social science. They explore e.g. the bridge between social and computational science, demonstrate theoretical approaches, discuss applicable methodologies, point out fields of activities, and illustrate limitations and restrictions in that field.

The second chapter »Data, Methods, and Instruments« deals with analytical and methodological approaches in the field of CSS. The authors

present different techniques and instruments to handle massive (online) data. They introduce us in e.g. (automated) data collection, tracking methods for analyzing social behavior, storage of large-scale (social) datasets, tools for collecting and storing historical financial data, potential benefits of using hierarchical as well as dynamic clustering approaches as well as information mapping and data visualization in the context of Big Data analytics.

The third chapter »Case Studies« puts theory into practice. The authors demonstrate the wide range of application on CSS theories, methods, and instruments. The studies explore the use of Big Data by examining e.g. human relationships and kinship, migration routes, and dynamics of protest, and the relevance for affective computing.

The fourth chapter »Tutorial Section« aims to introduce us in practical application of Computational Social Science. The authors demonstrate hands-on manuals in the field of learning analytics as well as social media monitoring.

References

ALVAREZ, R. (ed.): *Computational Social Science: Discovery and Prediction* (Analytical Methods for Social Research). Cambridge [Cambridge University Press] 2016

BOYD, D.; K. CRAWFORD: Critical Questions for Big Data. In: *Information, Communication & Society,* 15(5), 2012, pp. 662-679

BOYD, D. M.; N. B. ELLISON: Social Network Sites: Definition, History, and Scholarship. In: *Journal of Computer-Mediated Communication*, 13(1), 2007, pp. 210-230

CIOFFI-REVILLA, C.: *Introduction to Computational Social Science. Principles and Applications*. London [Springer] 2014

CONTE, R.; N. GILBERT; G. BONELLI; C. CIOFFI-REVILLA; G. DEFFUANT; J. KERTESZ; V. LORETO; S. MOAT; J. P. NADAL; A. SANCHEZ; A. NOWAK; A. FLACHE; M. SAN MIGUEL; D. HELBING: Manifesto of Computational Social Science. In: *The European Physical Journal Special Topics*, 214, 2012, pp. 325-346

JAHODA, M.; P. F. LAZARSFELD; H. ZEISEL: *Die Arbeitslosen von Marienthal. Ein soziographischer Versucht*. 25. Auflage, Frankfurt/M., Leipzig [Suhrkamp] 2015 [1933]

KING, G.: Preface: Big Data Is Not About the Data! In: ALVAREZ, R. MICHAEL (ed.): *Computational Social Science: Discovery and Prediction*. Cambridge [Cambridge University Press] 2016

LAZER, D.; A. PENTLAND; L. ADAMIC; S. ARAL; A. BARABÁSI; D. BREWER; N. CHRISTAKIS; N. CONTRACTOR; J. FOWLER; M. GUTMANN; T. JEBARA; G. KING; M. MACY; D. ROY; M. VAN ALSTYNE: Computational Social Science. In: *Science*, 323(5915), 2009, 721-723

SCHROEDER, R.; L. TAYLOR: Big Data and Wikipedia Research: Social Science Knowledge Across Disciplinary Divides. In: *Information, Communication & Society*, 18(9), 2015, 1039-1056

STEGBAUER, C.: *Wikipedia: Das Rätsel der Kooperation*. Wiesbaden [vs Verlag für Sozialwissenschaften] 2009