Computational Social Science: A Literature Review

Anne Suphan & Christopher Zirnig

@czirnig

Introduction

New data are being obtained, generated and analysed for very different purposes. The importance of big data, for example, for future social research is made clear in many writings. For example, Conte (2016) refers to Big Data as the promise of large and relatively cheap data sets as an opportunity for global data, focusing more on political decisions than on potential profits. Especially unstructured "big data" will be a major challenge for social science research methods. For example large amounts of texts can only be handled with advanced computational methods. Creating these methods will require the knowledge and theoretical background collected over the past 100 years in the qualitative social science approach.

Method

After a systematic database search on "scopus" with the search word "computational social science" we added 286 articles to our sample. From these 286 articles we drew a random sample of 53 articles that are being categorised and analysed closely, using content analysis. We will take a look at both established and quoted definitions of CSS and Big Data to try to reach a level of abstraction, that will allow us to make statements about the similarities and differences in the use of CSS and Big Data amongst the different scientific disciplines. We also analyse the abstracts of all 286 articles with a cluster analysis. We then decided to create a second dataset containing 857 articles that quoted Lazer et al. (2009) in order to get a better picture in which fields the term "computational social science" is used.

286 abstracts containing "Computational Social Science":

First Results

We determined five broad categories: The **first** category deals with computer science, neuroscience, biology and other sciences. It is about the development of artificial life and especially artificial intelligence (Saunders / Bown 2015 and Díaz / Domínguez 2013). CSS is used, for example, to enable machine learning (Hu et al. 2014 and Mason / Vaughan / Wallach 2014) and to better understand systems of multiple agents and man-machine communications (Conte / Paolucci 2014 and Youyou / Kosinskib / Stillwell 2015). The **second** category is about user-generated content from eg. social networks. This involves on the one hand methodological questions such as datelining (Olmedilla / Martínez-Torres / Toral 2016) and the development of new tools (Borra / Rieder 2014). Secondly, it also contains political issues such as political online events in China (Liu / Zhong / Chen 2016). In the **third** category we find forecasting and epidemiological inquiries such as Stohler (2014). The **fourth** category deals with the development of social science theory in the course of CSS developments and Big Data (eg. Shah / Cappella / Neuman 2015). The question is how social sciences are changing and may need to reposition as (a) discipline(s) in line with technical innovations. The **fifth** category is an "Others" category and summarises what is additionally discussed in the discourse about CSS. For example gender as part of CSS (Purohit u. A. 2016), Policy-Making (Lettieri 2016), new forms of public theories (Tufekci 2016) and urbanisation (Wei / Ye 2014).

Outlook

The definitions of CSS are broadly fanned and quite different and are still to be analysed and categorised to find a common denominator. Cioffi-Revilla, for example, argues (2014):

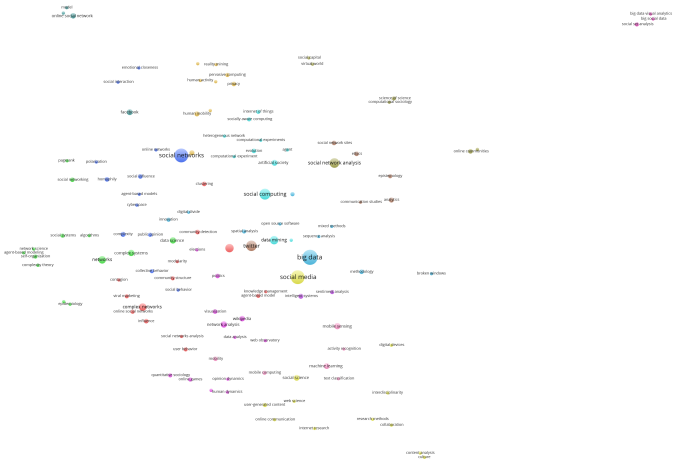
> *"The new field of Computational Social Science can be defined as the interdisciplinary investigation of the social universe on many scales, ranging from individual actors to the large largest groupings, through the medium of computation." (P. 2)*

Other studies define CSS as a better understanding of social behaviour (Strohmaier / Wagner 2014). Still others put instrument based CSS in line with microbiology, astronomy or nanoscience (Borra / Rieder 2014). **Our main goal** is to detect intersections among the field of CSS and visualise opportunities for fruitful interdisciplinary approaches. For example Hu et al. (2014) address the problem of dealing with large corpi of texts:

> *"Unlike information retrieval, where users know what they are looking for, sometimes users need to understand the high-level themes of a corpus and explore documents of interest." (P. 424)*

This quotation makes it obvious, that the theoretical background of Levi Strauss, Ervin Goffmann or Robert Park is highly needed in CSS methodology, dealing with unstructured text data. Computational Social Science is an interdisciplinary field and therefore cooperation is a necessity. We want to illustrate opportunities for cooperation. Topic modelling and qualitative social research is only the beginning.

857 abstracts that quoted Lazer et al. (2009):



References:

Borra, Erik/Rieder, Bernhard (2014): Programmed method: developing a toolset for capturing and analyzing tweets., in: Aslib Journal of Information Management 66, S. 262–278.

Cioffi-Revilla, Claudio A. (2014): Introduction to computational social science: principles and applications, London; New York: Springer.

Conte, Rosaria (2016): Big data: un'opportunità per le scienze sociali?, in: Sociologia e Ricerca Sociale, S. 18–27.

Conte, Rosaria/Paolucci, Mario (2014): On Agent-Based Modelling and Computational Social Science., in: Frontiers in Psychology 5, S. 1–16.

Díaz, Rodrigo Martínez/Domínguez, Álvaro Zapata (2013): Las ciencias sociales y los dispositivos de la complejidad., in: Cuadernos de Administración 29, S. 123–131.

Hu, Yuening/Boyd-Graber, Jordan/Satinoff, Brianna/Smith, Alison (2014): Interactive topic modeling., in: Machine Learning 95, S. 423–469.

Lazer et al. (2009) Computational Social Science. Science Vol. 323, Issue 5915, pp. 721-723.

Lettieri, Nicola (2016): Computational Social Science, the Evolution of Policy Design and Rule Making in Smart Societies., in: Future Internet 8, S. 1–17.

Liu, Tieying/Zhong, Yang/Chen, Kai (2016): Interdisciplinary study on popularity prediction of social classified hot online events in China, in: Telematics and Informatics.

Mason, Winter/Vaughan, Jennifer Wortman/Wallach, Hanna (2014): Computational social science and social computing., in: Machine Learning 95, S. 257–260.

Olmedilla, M./Martinez-Torres, M.R./Toral, S.L. (2016): Harvesting Big Data in social science: A methodological approach for collecting online user-generated content, in: Computer Standards & Interfaces 46, S. 79–87.

Purohit, Hemant/Banerjee, Tanvi/Hampton, Andrew/Shalin, Valerie L./Bhandutia, Nayanesh/Sheth, Amit (2016): Gender-based violence in 140 characters or fewer: A #BigData case study of Twitter., in: First Monday 21, S. 1–1.

Saunders, Rob/Bown, Oliver (2015): Computational Social Creativity., in: Artificial Life 21, S. 366–378.

Shah, Dhavan V/Cappella, Joseph N./Neuman, W. Russell (2015): Big data, digital media, and computational social science: Possibilities and perils., in: Annals of the American Academy of Political and Social Science 659, S. 6–13.

Stohler, Christian (2014): The Next Frontier: Digital Disease Detection in Cyberspace., in: Journal of Oral & Facial Pain & Headache 28, S. 105–105.

Strohmaier, Markus/Wagner, Claudia (2014): Computational Social Science for the World Wide Web., in: IEEE Intelligent Systems 29, S. 84–88.

Tufekci, Zeynep (2016): As the Pirates Become CEOs: The Closing of the Open Internet., in: Daedalus 145, S. 65–78.

Wei, Yehua Dennis/Ye, Xinyue (2014): Urbanization, land use, and sustainable development in China, in: Stochastic Environmental Research and Risk Assessment (SERRA) 28, S. 755–755.

Youyou, Wu/Kosinskib, Michal/Stillwell, David (2015): Computer-based personality judgments are more accurate than those made by humans, in: Proceedings of the National Academy of Sciences of the United States 112.