Big data and digital methods in science communication research: opportunities, challenges and limits

Nico Pitrelli

Abstract

Computational social science represents an interdisciplinary approach to the study of reality based on advanced computer tools. From economics to political science, from journalism to sociology, digital approaches and techniques for the analysis and management of large quantities of data have now been adopted in several disciplines.
The papers in this *JCOM* commentary focus on the use of such approaches and techniques in the research on science communication. As the papers point out, the most significant advantages of a computational approach in this sector include the chance to open up a range of new research opportunities: from the study of technical and scientific controversies to citizen science, from the definition of new norms and practices for science journalism to open science issues.
On the other hand, difficulties are shared with other areas of application. The main risk is that the large quantity of data available can overwhelm the importance of theory. Instead, as the papers in this commentary demonstrate, big data should push scientists to pursue a deeper epistemological and methodological reflection also in the research on science communication.

Keywords

Science and media; Science communication: theory and models; Visual communication

We now live life in the network, and this represents a turning point for social research. This is the concept expressed by the authors of an article published in *Science* in 2009 [Lazer et al., 2009], now considered a reference in the fledgling history of computational social science: the huge availability of data and the digital traces we leave in our Facebook and Twitter discussions, in our regular exchanges of e-mails and texts, when we swipe our credit cards for our purchases, and in the images captured by video cameras are at the core of a dramatic change in epistemological and methodological terms. We are witnessing the emergence of a new disciplinary field — as Lazer et al. highlighted in their article — on which sociology, statistics, computer science, mathematics, economics and political science converge. Hence, according to the authors, big data and computational analysis techniques should become the bread and butter of social scientists.

The appeal launched from the pages of *Science* has not gone unheeded. In the second edition of their collection of papers, recently published by Sage, Nigel G. Fielding, Raymond M. Lee and Grant Blank — editors of one of the most

internationally-successful online social research manuals — attested to the significant growth, over slightly less than a decade, in the use of digital research techniques by social scientists from various disciplines [Fielding, Lee and Blank, 2017]. The birth of academic societies and specialised journals, a number of international conferences, the publishing of books for the general reader and, last but not least, the professionalisation of data scientists all confirm the prominence acquired by computational social science over the past few years.

In fact, automated data mining, social network analysis applied to the web, computational models used to understand how people interact, augmented reality, geospatial analysis are gaining ground every day more in multiple areas of application, ranging from the study of economic inequalities to education, from democracy to healthcare.

The papers featured in this *JCOM* commentary were written with a view to understanding whether and how science communication is part of the list of disciplines affected by the emergence of digital methods.

As Federico Neresini points out in his paper, this is a particularly significant issue in the context of PCST (*Public Communication of Science and Technology*) because, whilst the challenges and the opportunities connected with big data and computational techniques concern sociology in general, this is even more true for a subject of study — science and technology — which is very much present in the digital communication flow.

The research projects presented here have a background assumption that is similar to other works about the chances to innovate social research with "the emergence and the normalisation of web 2.0" considering that the net is "a place belonging to society and it should be conceived as such, rather than as a space outside society" [Boccia Artieri, 2015b].

For example, in his analysis of the communication strategies related to the anti-vaccine documentary *Vaxxed*, Davide Bennato highlights the principle according to which social media do not simply provide a new communication channel, but they truly represent an actual social space. Starting from this perspective, Bennato aims to demonstrate that, within scientific controversies, social platforms may promote dynamics that are typical of public relations rather than of the *Public Understanding of Science*. Its conclusions aside, the paper is interesting for the methods used to reach its results, which pertain to the field of *Search Engine Optimization*, as well as to the methods used to write and edit Wikipedia entries.

While Bennato describes some of the most interesting approaches within computational social science in order to analyse controversies in digital spaces, in his work Neresini illustrates the great opportunities offered by the digitisation of traditional media. His paper describes the TIPS (Technoscientific Issues in the Public Sphere) project, a platform able to automatically collect and organise a large quantity of digital contents available online mainly taken from the news published in the major Italian, British and French newspapers. One of the TIPS project goals is to analyse the media discourse on science and technology, having at disposal an amount of articles that was unthinkable up to a few years ago. Neresini also

describes the epistemological and methodological limits related to the use of analysis tools for large quantities of data. For example, one of the risks is treating automatic processing software as a "black box" able to replace the judgment of human researchers.

In his paper focusing on the computational instruments for text analysis, Yurij Castelfranchi also reflects on the promises and the issues arising out of the availability of huge databases. Castelfranchi presents some of their applications that are relevant in the field of social studies on science and science communication. The Italian-Brazilian researcher takes computer-assisted text analysis techniques as a starting point to demonstrate how increasingly blurred the boundary between qualitative and quantitative methods is becoming and to critically discuss the — possibly too simplistic — forecasts on the "end of theory" in the big data era.

The subject of the paper by Stuart Allan and Joanna Redden is the challenges science journalism will have to face in order to deal with the extraordinary quantity of data collected within citizen science projects. The remarks by the two scholars address issues related to privacy, web disinformation, the opacity of algorithms, the quality of data, which meet the epistemological needs of citizen science. Science journalists are required to devise new information strategies to contextualise and broaden the debate on the risks and opportunities arising out of the production of large datasets by citizens.

The commentary ends with the paper by Cristina Rigutto on online visual communication of science. Again, she addresses the role played by non-expert audiences in the co-production of scientific images spread on various web platforms. Rigutto maintains that, in order to identify interpretation patterns across the millions of pictures uploaded on the web every day — aside from considering the participatory nature of the editing process involving pictures, graphs, scientific drawings in different digital contexts — any research strategy on online visual communication of science should take into account the specific characteristics of the various social media and the dynamics producing and spreading information across online communities.

Globally, the picture outlined in the papers in this collection shares a few elements with the general reflection on the use of computational techniques in social research and introduces other elements specific to science communication. For example, it clearly highlights the interdisciplinary nature of computational social science, which is to be considered — as other scientists already pointed out [Boccia Artieri, 2015c], — as a meeting and experimentation place for researchers with different backgrounds rather than an independent disciplinary field. Another aspect emerged in similar remarks is the suggestion to take a balanced approach to big data. In fact, the awareness on the great development potential they offer should be matched by a constant epistemological and methodological reflection on the limitations of such big data being used without the support of relevant research purposes. As concerns science communication in particular, new and promising opportunities apparently lie ahead, from the study of technoscientific controversies to citizen science, from the definition of new norms and practices for science journalism to open science issues.

In this commentary we tried to provide a relevant representation of the opportunities of research in the field of science communication using computational techniques. We hope that the prospects illustrated may function as an incentive to broaden the interest in the approaches described, noting that while the big data revolution poses a cultural, social and technological challenge that scholars studying scientific knowledge spreading and learning processes will have to face, it also requires the adoption of new conceptual viewpoints, at the epistemological and methodological level, as well as in terms of skills that are needed.

Translated by Massimo Caregnato

References

Boccia Artieri, G., ed. (2015a). Gli effetti sociali del web. Forme della comunicazione e metodologie della ricerca online. Milano, Italia: Franco Angeli.

— (2015b). 'Gli effetti sociali del web: introduzione'. In: Gli effetti sociali del web. Forme della comunicazione e metodologie della ricerca online. Ed. by G. Boccia Artieri. Milano, Italia: Franco Angeli, p. 11.

— (2015c). 'Teoria e metodologia per la ricerca sul web sociale: tra Big Data e Deep Data'. In: Gli effetti sociali del web. Forme della comunicazione e metodologie della ricerca online. Ed. by G. Boccia Artieri. Milano, Italia: Franco Angeli, p. 42.

Fielding, N. G., Lee, R. M. and Blank, G., eds. (2017). The SAGE Handbook of Online Research Methods. SAGE Publications Ltd.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. (2009). 'Social Science: Computational Social Science'. *Science* 323 (5915), pp. 721–723. DOI: 10.1126/science.1167742.

Author

Nico Pitrelli, Ph.D., is the co-director of the "Franco Prattico" Master's Degree in Science Communication at the International School for Advanced Studies — SISSA in Trieste. His research focuses on the training in science communication, scientific journalism and the impact of the net on the production and circulation of scientific knowledge. E-mail: pitrelli@sissa.it.

How to cite

Pitrelli, N. (2017). 'Big data and digital methods in science communication research: opportunities, challenges and limits'. *JCOM* 16 (02), C01_en.

ISSN 1824-2049. Published by SISSA Medialab. jcom.sissa.it