Computational research in media studies:
Methodological implications

Julia Noordegraaf*

Media studies as an academic discipline within the humanities has existed for about 40 years and focuses on the content, history, and effects of various media, in particular mass media such as newspapers, film, radio, television, and Internet-based social media platforms. Methodologically, the discipline has been strongly influenced by the critical theory of the Frankfurt School, which targeted the alienating effects of market-driven mass media as well as their leveling effect on the arts and other forms of ‘high culture’. In the 1970s, the rise of cultural studies introduced reflections on class, gender, and race. More recently, an interpretivist ‘social constructivism’ has become the dominant theoretical approach, arguing that ‘truths’ are socially constructed (Schroeder, 2014: 173). Media studies research is often small-scale (focused on one film or television series, or on one specific place or time period) and characterized by a hermeneutic, interpretative analysis of media content and/or a critical historical analysis of the social and economic contexts of its production, distribution, and reception.

In the context of large-scale digitization of cultural heritage over the past decade, a growing amount of historical media-related sources has become available in digital form. Besides, the emergence of the Internet and of affordable digital technologies for recording video and audio has led to rapidly expanding collections of new audiovisual media content produced by amateurs and disseminated via social media platforms such as YouTube, Vimeo, Instagram, Twitter, and Facebook. In addition, media scholars have increasingly adopted software available for the creation of databases with structured data on various aspects of production, distribution, and reception contexts (Ross et al., 2009: 8). In parallel, various new tools have been developed for searching these new, digital collections, and analyzing the data contained in them, such as tools for text mining, image analysis, geographical mapping, or network visualization (Ross et al., 2009). The creation of these tools has been primarily driven by computer scientists interested in developing tools with generic search, analysis, and/or visualization functionalities that can be applied to a number of different collections and research questions.

At present, a gap exists between the affordances of digital data and computational tools, and their application in media studies, in spite of the growing number of successful experiments in this area. In this essay, I argue that the, as of yet, modest uptake in mainstream media studies can be explained by the fact that introducing computational research requires scholars to review and complement the methods

* Prof. dr. Julia Noordegraaf is a media scientist and works as professor of digital heritage at the Department of Media Studies of the Faculty of Humanities, University of Amsterdam. Email: j.j.noordegraaf@uva.nl.
they traditionally work with. Moreover, as it stands now the tools often lack the transparency that humanities scholars require for interpreting the results, or they insufficiently engage with the cultural criticism that characterizes media studies research. At the same time, I argue, there is a real need to engage with computational research because, first, scholars are already making use of search tools and databases, often without critically reflecting on methodological implications. Second, a lack of engagement with digital sources and tools will put these scholars at a disadvantage in a world in which social and cultural life is becoming increasingly digitized (Graham et al., 2015: 33).

Methods in media studies

In her introduction to methods of media and cultural studies, Jane Stokes (2003) identifies three strands of research, focused on: first, industry (the production and distribution of media content); second, text (the style, structure, and meaning of media forms); and third, audience (reception of media content). However, in practice one mostly observes a binary distinction between, on the one hand, text-oriented scholarship (interpretative, critical analysis of media texts) and, on the other, context-oriented scholarship (analyzing the socio-economic networks of production, distribution, and reception).

The relatively young discipline of media studies is characterized by a fundamental methodological hybridity (Bod, 2010: 420). Researchers use a great variety of methods that are often combined in a single study. Bonnie Brennen (2012) identifies six methods used in media studies research: interviews, focus groups, history, oral history, ethnography and participant observation, and textual analysis. Based on an explorative study of articles published in two major cinema and media studies journals between 2010 and 2015, it appears that in the humanities branch of media studies, in particular in the fields of film and television studies, the most prominent methods are textual analysis (close reading of media texts and practices, including narrative analysis), discourse analysis, history and, to a lesser extent, interviews and oral history.3

Media researchers usually work with existing datasets: audiovisual sources (film, radio and television programs, online audio and video); textual sources (primary and secondary literature, newspapers, TV guides, social media content); and structured data (such as TV audience ratings, or databases such as Cinema Context4 with data about films, cinemas, screenings, and people and companies active in the film sector, taken from newspapers and archival sources). Recent user studies (Bron et al., 2013; Bron et al., 2015) have shown that media researchers generally analyse various sources in relation to each other. A good understanding of audiovisual sources entails detailed knowledge of the contexts in which they were produced, circulated, and viewed (Maidment, 2012: 113). Researchers working with audiovisual sources are often also interested in the role of the media in the construction, circulation, and reception of a specific social or cultural phenomenon. This requires that various perspectives on the same phenomenon can be traced through time and analyzed in relation to each other. Moreover, the research process for media re-
searchers is characterized by an explorative phase on the basis of the content of material (in which the research question is developed and/or fine-tuned) followed by a more focused collection and analysis of, mainly, contextual data (contextualization phase) – a cycle that is often repeated once or several times, and is rounded off with a presentation of the results (presentation phase) (Bron et al., 2015). The specific nature of audiovisual sources partly explains why media studies scholarship is generally small-scale and focused on a limited corpus and/or a limited historical period. Audiovisual media are time-based and need to be watched and listened to in order to grasp their content. Often programs are not tagged at the level of scenes and shots, so entire programs have to be viewed before being able to analyze specific sections. In addition, because of copyright constraints and lack of digitization, access to media-related content is often limited and requires on-site visits to media archives, which is time-consuming and costly (Bron et al., 2015).

State of the art of computational research in media studies

In recent years, a number of efforts have been made to facilitate access to media-related sources. Large-scale digitization projects by libraries, archives, and museums have generated digital collections of newspapers, radio and television programs, films, and related content, such as media periodicals and program guides. The same institutions have also created platforms to provide direct access to these collections. In addition, media scholars have collaborated with computer scientists and heritage partners to develop tools for searching, analyzing and visualizing these digitized collections.

At the same time, the availability of affordable and readily available software has stimulated researchers to set up their own databases, often with structured data relating to the context of media production, distribution, and consumption. As Ross et al. (2009: 8) note, the creation of these databases as well as the opportunities for collaboration provided by the Internet have entailed an increasing interest in empirical research in media studies. Current research that uses digital data and computational tools for search, analysis, and visualization of media-related content falls into two domains. The first domain comprises research centered on textual analysis of media content. Examples include the online collaborative platform Cinemetrics, developed by film scholar Yuri Tsivian, which allows for automated analysis of film style based on the statistical calculation of average shot lengths; and the Austrian Digital Formalism project that employed visual analysis software for the computational understanding of media aesthetics in the work of the Russian avant-garde filmmaker Dziga Vertov. Here, the use of software can enable the detection of structural and stylistic features that cannot be perceived with the human eye or are too labor-intensive to track without computational support. The second domain comprises social and economic historical research into, in particular, the distribution and reception of media content. Examples include studies into the social and geographical spread of specific film or television genres (e.g. Boter & Pafort-Overduin, 2009; Olesen et al., 2015), or measuring the popularity or success of a medium, genre, or single title or
star (e.g. Sedgwick, 2009; Dibbets, 2010; Van Beusekom, 2013). As has become clear from the first experiments in this domain, there is potential in advanced statistical analysis of historical data on the media industry, in that it can significantly increase both the scope (geographically, temporally) and complexity of the research (combining multiple dimensions, such as location, genre, and social and cultural characteristics of audiences). At the same time, as Ross et al. (2009: 11) conclude, these examples also show that this type of research requires a fair amount of statistical experience – an experience most ‘traditional’ media scholars do not possess. Besides this methodological challenge, there is a number of other reasons that explain why the potential of digital data and tools in media studies research has not yet been fully exploited.

Methodological implications

First, notwithstanding the great achievements of heritage institutions and scholars in creating digital collections, the availability of comprehensive, reliable datasets is still rather limited, especially when it comes to the earlier periods of media history. Another problem is that even datasets on the same phenomenon (e.g. cinema culture) are not always based on the same data model, complicating their combined use. As Ross et al. point out (2009: 9), international collaboration between scholars and heritage partners based on a shared data collecting model might be the road towards creating more comprehensive, interoperable datasets. At the same time, the lack of comprehensive collections does not preclude data-driven research; as Graham et al. (2015: 25) emphasize, what counts as ’big data’ in digital historiography is in the eye of the beholder, and concerns any collection of materials that is too large to be read or viewed in a reasonable amount of time or that requires computational intervention to make sense of it.

A second obstacle to a broad uptake of digital collections in media studies research is the lack of access to audiovisual sources and related materials (Bron et al., 2015). Even in institutions involved in large-scale digitization projects, copyright restrictions limit a wide availability of audiovisual sources and published materials. A solution to this problem is to create research infrastructures that include user identity management services for controlling access to copyright protected materials, as is the objective of the Dutch CLARIAH project.10

In addition to the scope and accessibility of sources, a third challenge to the uptake of digital data and tools in media studies research is the required knowledge of computational technology and of different research methods, in particular statistics. Knowledge on the underlying principles of tools for searching, analyzing, and visualizing large data collections is required for the correct interpretation of results. Media scholars are generally trained in interpretative, qualitative methods of analysis and rarely use software designed to support such analysis, nor are they familiar with the algorithmic logic underlying such software. Moreover, since the scope of ‘Big Data’ by definition exceeds the capacity of a single human interpreter to make sense of them, research based on large datasets by necessity involves some statistical operation or the other (e.g. counting word frequency relative to other
words for creating graphs or word clouds). This feature of data-driven research demands a basic level of familiarity with the principles of statistics which most media scholars currently do not possess. On the technology side, many computational tools for search, analysis, and visualization provide attractive, user-friendly interfaces that ‘black box’ underlying statistical operations, making them less transparent for researchers.

Although the choice of method always depends on the specific problem studied, a basic knowledge of database architecture and the algorithms behind search, analysis, and visualization software is indispensable for all humanities scholars, if only because they already work with databases such as JSTOR and Google Books and with search engines such as Google Search. In order to equip humanities scholars for an increasingly digital future, it is essential to set up training programs that familiarize them with databases and software for qualitative and quantitative analysis. In addition, tool builders have to increase transparency by providing guidelines clarifying the provenance and scope of underlying data and the statistical operations performed on them.

Having said that, the level of penetration of software-supported research in media studies also has its limits in that, at least at present, most computational tools do not facilitate a critical engagement with culture (Liu, 2012; Nyhan, Flinn & Welsh, 2015). The critical strand in media studies research regards the new interest in empirical research and the move towards quantitative and computational research as an attempt to shift humanities research in the direction of positivist, empirical scientific approaches (Schroeder, 2014: 165). Many scholars in the humanities that adopt a critical analysis approach demonstrate a deep resentment to computation as a cultural phenomenon associated with neoliberalist tendencies in contemporary society (e.g. Golumbia, 2009). And even scholars who are interested in the affordances of computational research point to the fact that the semantic complexity of culture often requires an approach that exceeds the capacities of automatic analysis (Ross et al., 2009: 15). However, these comments do not preclude software-supported research, whether qualitative or quantitative, or, a combination of both. Naturally, there is room for different approaches, and a software-supported analysis does not prevent researchers from also conducting more traditional, analog ones. As Graham Gibbs warns, we should not confuse analysis software with a method, but rather see it as ‘organizational support’ for a specific, predefined method of analysis, whichever method that may be (2013: 291). The challenge is to experiment with existing tools and define requirements to make them more responsive to the specific needs of the researcher.

Perhaps the greatest contribution of computational research in media studies is that it invites a clear articulation of the theoretical approach, the chosen method, and the selection of sources and means of analysis. The transparency that scholars require of digital datasets and analysis software thus also reflects back on the way they reflect on their own research practice and methodology (see Weller, 2012: 16). This not only concerns research based on software for quantitative, statistical analysis but also research driven by qualitative data analysis, where software can support the articulation of the conceptualization and interpretation process that takes place in the mind of the researcher and that otherwise remains implicit (e.g. Kloep-
fer, 2009). The examples of computational research that currently exist demonstrate the potential of broadening the scope and complexity of media studies research. Besides, the rapid growth of collections of source material available in digital form demands the use of software to manage and make sense of them. If media scholars invest in the digital literacy and methodological awareness required for critically working with digital data and tools, and collaborate with computer scientists and heritage institutions to increase the transparency of analysis tools, the field of media studies will be fully equipped for the digital twenty-first century.

Notes

1 This article has greatly benefited from the feedback provided on the first draft by José van Dijck, Jasmijn van Gorp, Thomas Poell and the anonymous reviewer, for which I kindly thank them.

2 Although the study of the social and cultural impact of mass communication technologies such as photography, newspapers and film has a much longer history, the origin of media studies as a separate academic discipline may be marked by the foundation of the first academic graduate program in Media Studies in the United States at The New School in New York in 1975, https://en.wikipedia.org/wiki/History_of_media_studies (accessed 17 November 2015). In the present essay, I limit my discussion to the humanities branch of media studies that focuses on the cultural dimensions of media practices and content, ignoring the field of communication studies as practiced in the social sciences, which generally has a different theoretical and methodological orientation. I also particularly focus on developments in the major sub-disciplines of film and television studies, because the newer field of digital media and Internet studies has its own dynamics regarding computational methods and is often crossing over into the social sciences.

3 In order to assess the most prominent research methods in current media studies, I conducted an exploratory, small-scale, manual analysis of the methods employed in the feature articles that appeared between 2010 and 2015 in two major journals. A review of the feature articles in Cinema Journal (the journal of the Society for Cinema and Media Studies, the largest professional association for film and television scholars) between 2010 and 2015 (vol. 49:2 to vol. 55:1, in total 24 issues with 130 feature articles) reveals that the majority of articles is based on textual analysis (69%, n=89), with a smaller number of articles focused on historical research on context (23%, with 18% (n=24) focused on industry and 5% (n=6) on audience/reception), and a small category of articles with a more theoretical or philosophical orientation (8%, n=11). The contributions to NECSUS: European Journal of Media Studies (the journal of the European Network for Cinema and Media Studies) often combine different qualitative methods, in particular textual analysis or historical research combined with discourse analysis and/or critical theoretical or philosophical reflection. In the seven issues that have appeared since 2012, only three out of 81 articles combine qualitative methods with quantitative ones, all three in the area of digital media studies.

4 www.cinemacontext.nl.

5 For example, Dutch archives in the Images for the Future Program digitized about 140,000 hours of film and video, 310,000 hours of audio and 2.5 million photographs. “Beelden van het verleden: 7 jaar beelden voor de toekomst” (final report Images for the Future project, Nederlands Instituut voor Beeld en Geluid, EYE Filmmuseum, Nationaal
In the Netherlands, these include the tools for media-related research that are currently sustained and made available to researchers via CLARIAH: the Common Lab Research Infrastructure for the Arts and Humanities (http://www.clariah.nl/projecten/focusgebieden/media-studies/255-wp5).

For example, in the field of cinema history these include the filmography and newspaper-based databases of the Canadian GRAPHICS project (Pelletier & Véronneau, 2009), the London Project with data on the London film business (http://londonfilm.bbk.ac.uk/), the Cinema Context database on Dutch film culture (http://cinemacontext.nl/), the Australian Kinomatics database on international film culture (http://kinomatics.com/) and data on Belgian cinema culture collected within the Verlichte Stad (Illuminated City) project at the universities of Antwerpen and Ghent (Biltereyst & Meers, 2014).


References


Bron, M., Van Gorp, J. & De Rijke, M. (2015). Media studies research in the data-driven age:


