1. A Doubtful Handshake

The Promising Frontier of Data-Driven Film and Media Studies

Giorgio Avezzù and Marta Rocchi

Charles Marion Russell was a prolific American artist who, in the early 1900s and the first decades of the new century, created powerful paintings depicting the wild west of the United States. These paintings also had some influence on the cinematic imagery of the western genre. One of his beautiful watercolours from 1910, titled "A Doubtful Handshake", shows a cowboy and a Native American leaning toward each other, each on their own horse, to shake hands, albeit with some suspicion on both sides, against the backdrop of a desolate frontier landscape and snow-capped mountains under an off-white sky. The artist later returned to the same subject, and a similar oil painting is part of the British Royal Collection under the title "Left Handshake Is the Safest" – the right hand is ready with a rifle, just in case.

It seems to us that the image of the doubtful handshake is an appropriate way to represent the encounter that we also hope to promote with this volume, as well as the difficulties that may arise, such as mistrust and resistance. We are referring to the intersection between film and media studies, on the one hand, and the culture of data and data-driven analysis tools, on the other. In other words, we are ultimately talking about a particular opportunity to bring together the "two cultures" (Snow 1959), given that studies on cinema and audiovisual media mostly come from the humanities tradition, while it is generally the STEM disciplinary fields that deal with data. We do not want to force the metaphor too much, and establish who are the cowboys and who are the Indians, and who keeps his right hand on the rifle, but certainly the encounter we are referring to represents a frontier theme, which has become essential today as a result of the innovations introduced by the digital revolution in our field of study, in the realm of the objects that we are used study.

Certainly, it is an issue that is impossible to ignore today, and one that we feel is urgent to confront, even more so now than in the past. However, it is not a theme that originated with the digital era: one could attempt a sort of archaeology of the approaches that have tried to base media studies on the most solid foundations of the so-called hard sciences. Among these, attempts to "datify" the textual aspects of films for their formal and stylometric analysis are well known. These are attempts that had already been anticipated by some previous experiments, but that began to gain some diffusion from the mid-1970s. Think of the analysis of average shot length to investigate the stylistic signatures of films directed by different directors, or to survey cutting rates in different periods of film history (e.g., the 1940s and 1950s) (Salt 1974, 1983, Tsivian 2009, Heftberger 2018, see also Bordwell 2006). Similar quantitative analyses, as well as those on the scale of shot and camera movement, have been carried out for decades, literally manually, without recourse to automated tools.

The strength of objectivity of the quantitative data of "cinemetrics", its apparent self-evidence, has perhaps sometimes led to sacrificing the depth of interpretation of the data itself. This has also been observed recently by Franco Moretti – a scholar who cannot be suspected of being hostile to a "distant reading" approach – precisely in relation to certain perhaps hasty conclusions on alleged trends observable from data that would show the shortening of the duration of shots in American cinema over the decades (Moretti 2022, about Cutting and Candan 2015). It seems to us that even this kind of objection basically interprets the old opposition between the "two cultures" and the difficulties of communication between one and the other.

The contrast between more traditional approaches and data-driven analysis in cinema studies, on the other hand, is no doubt well understood by anyone who reads the pages of a paradigmatic scholar like Barry Salt (1983), particularly the first chapters of his *Film Style and Technology*, "the first and only history of motion picture style". One cannot help but notice in those pages a certain arrogance on the part of the author, who boasts a PhD in theoretical physics and believes in the objectivity of what he calls "real sciences", and an explicit aversion to the humanities in general and their alleged charlatanism. An aversion that was motivated, for that matter, also by the hostility reserved for Salt himself by many "traditional" film scholars as well as by all those major film journals that denied him, "with no reason given", the publication of earlier versions of his studies. These same tensions seem to us to often continue to persist even today, but perhaps it

would be a case of finally tempering or dissolving them – at least this is what we would like to contribute to doing – by reconciling the scientific reliability of the method and the theoretical fecundity of the research question, and harmonizing the objectivity of analysis with the depth of interpretation. As Charles Percy Snow (1959) himself said, "This polarization is sheer loss to us all. To us as people, and to our society".

This book focuses on the datafication of audiovisual media content and audiences, as well as data-driven methods and methodologies for studying films and TV series. While quantitative perspectives applied to narrative audiovisual texts and audience behavior have existed in the history of film and media studies, as noted above, they have often been met with resistance from proponents of more traditional approaches. However, there is an increasingly established tradition of quantitative studies at the international level. Influential researchers have emphasized the potential of these studies, as well as the gap between the possibilities offered by digital data and computational tools, and their actual application in the field of media studies (Noordegraaf 2016). In this context, various methodological approaches have been developed to uncover patterns, trends, and characteristics of audiovisual products. These approaches include processing textual objects through multimodal techniques, analyzing social discursivity using automated software, and evaluating production and consumption processes through social network analysis and modeling techniques. Each of these research projects is grounded in specific tools for collecting and analyzing audiovisual media data, using statistical and computational techniques. The aim of this volume is to draw attention to and facilitate discussion on systematic methods for conducting data-driven research in film and media studies. Specifically, it delves into the insights that data-driven approaches can offer for the study of contemporary audiovisual products. We share the belief that "a unification of the field might come about not in the first instance, through a standardization of the elementary research procedures, but in the final instance, through an open-ended application and comparison of multiple methodology" (Jensen 2020).

The volume features some of the contributions from the program of the thirteenth edition of the Media Mutations conference, titled *Audiovisual Data: Data-Driven Perspectives for Media Studies*, which was held in Bologna on October 6th and 7th, 2022. The conference, as well as this book, was intended as an opportunity both to investigate *what we can do with data* and the analytical and interpretative possibilities at stake, and to reflect on

what data can be (and what kind of data we must deal with), in order to problematize the possible limits of such approaches to the study of audiovisual narrative media. To gain a better understanding of what can be done with data, and the potential of data itself, we will briefly explore some of the most popular and promising areas of research in media studies. This will involve considering several interesting methods and approaches to data-driven research that are relevant to the contemporary landscape of media studies.

One of the key areas of inquiry related to data-driven research in media studies is media content analysis. This method has a rich history, which Macnamara (2018) details in his work. Initially, media content analysis was developed as a systematic method to study mass media, specifically propaganda (Lasswell 1927). Over time, content analysis became an important research method for studying portrayals of violence, racism, and representations of women in television programming and films. This is because such symbolic content has the potential to influence attitudes and behavior. These approaches are still prevalent today. For example, Greenberg and colleagues (2009) compared food and drink consumption across different television program genres using content analysis. Himes and Thompson (2017) analyzed the specific gender, age, and types of fat stigmatization frequently depicted in movies and television shows using content analysis. Chapoton and colleagues (2020) utilized content analysis to identify patterns and frequencies of occurrences linked to alcohol within TV series. Content analysis has also been implemented to examine depictions of gender (Sink and Mastro 2017, González-de-Garay et al. 2020) and immigration (Ramos et al. 2019) on primetime television and to investigate narrative features in US medical TV series (Rocchi and Pescatore 2022). By examining media content, researchers can gain valuable insights into how media might shape attitudes and behaviors related to important social issues.

Despite its long history, content analysis continues to be a valuable tool in media studies research, as it enables researchers to make replicable and valid inferences by interpreting and coding audiovisual data, and converting qualitative data into quantitative data. In this context, computer-assisted analysis has significantly improved the process of content analysis, allowing researchers to boost the conversion of qualitative to quantitative data. Moreover, with the advent of machine learning techniques, such as topic modeling and natural language processing, data-driven studies in several areas of interest for media studies have become possible. For instance, studies can recognize characters from scripts, dialogues, captioned images,

and video material (Cour et al. 2010, Bojanowski et al. 2013, Parkhi et al. 2015, Huang et al. 2020, Nagrani and Zisserman 2018), conduct linguistic analysis on TV series dialogue (Bednarek 2018, 2020, Tukachinsky et al. 2022), or create screenplay summaries (Gorinski et al. 2015, Papalampidi et al. 2020).

An interesting perspective of media content analysis is its intersection with the study of diversity issues. Indeed, by conducting qualitative and quantitative content analysis, it is possible to study on-screen diversity (Sink and Mastro 2017, Ramos et al. 2019, González-de-Garay et al. 2020) and relate it to one of the components of off-screen diversity, which is the diversity of the workforce within individual media outlets, also known as "source diversity" (Napoli 1999). Quantitative investigations have already explored who is responsible for on-screen representation within the production context. For instance, Loist and Prommer (2019) conducted data collection on different sources, such as funding and box office, and empirically highlighted how gender imbalances in the German film industry contribute to the unsustainability and precarity of women creative professionals' careers, such as directors, producers, cinematographers, and scriptwriters. To monitor and assess large-scale off-screen diversity within the audiovisual production industry, one of the most useful tools is to use IMDb as a data source (Fanchi and Tarantino 2019, 2020, Verhoeven et al. 2019, 2020), although this approach has some limitations. Professional data, product information, and production details can also be useful for exploring broader phenomena that are usually difficult to investigate in their totality, using Social Network Analysis (SNA) techniques. For instance, Verhoeven and colleagues (2020) applied network science to the question of gender inequality within the Australian, German, and Swedish film industries, and showed that the expansion of collaborative sector networks between men and women is the most likely way to generate behavioral changes that improve network openness and address gender inequality. SNA techniques have also been useful in investigating the production of US TV series, between capital and labour strategies (Ruffino and Brembilla 2016) and cinema of "cultural interest" (Holdaway 2016). Furthermore, the interaction among audiovisual characters is a growing area of interest and can be analyzed in relation to textual analysis to predict the performance of TV series (Fronzetti Colladon and Naldi 2019, Hunter 2019).

Artificial intelligence (AI) is certainly a crucial area of investigation in data-driven research in media studies. Indeed, with the increasing signifi-

cance of data analytics and artificial intelligence strategies, many creative industries are now leveraging these technologies to improve their production processes (Anantrasirichai and Bull 2022). As a result, the number of data-driven studies in film, media, and sound studies has also grown substantially. Artificial intelligence is now being used for a range of tasks, including voice, text and image recognition (Fischer et al. 2011, Doukhan et al. 2018, Nagrani and Zisserman 2018), archives practices (Colavizza et al. 2021, De Quintana et al. 2021), decision-making support in the film industry (Ghiassi et al. 2015), and content-based predictive models of consumer behavior (Toubia et al. 2019, 2021). With the increasing use of media data, AI is being employed in various areas for the production and automatic generation of narrative audiovisual content, such as video editing, clip selection, filming, and screenplay writing. For instance, Smith and colleagues (2017) used AI to create a real horror movie trailer by modeling the key defining components that characterize horror movie trailers, while Bost and colleagues (2019) developed a tool for the automatic generation of video summaries of TV series that models the dynamics of the plot and extracts relevant sequences to create character-oriented video summaries. Derda (2022) has recently identified a new trend in the audiovisual production process, where artificial intelligence is now being used to support all phases of production. Traditionally, audiovisual production has been composed of three main phases: pre-production, production, and post-production. However, with AI-Supported Production Process, the incorporation of evaluations and feedback from the collection, analysis, and application of consumer data can now be integrated into all phases of audiovisual production, from initial conceptualization to final distribution. Derda (2022: 12) explains that this approach involves a "repetitive, or even continuous, evaluation process of the performance of the content and analysis of audience feedback", resulting in a shift in the nature of the production process from a relatively linear to a cyclical one. Artificial intelligence is revolutionizing the creative industries by enabling producers to gain valuable insights into consumer behavior and preferences, which can inform their creative strategies and deliver content that resonates more effectively with audiences.

The intersection between data analytics and theory has been explicitly sought in research that addresses audiovisual media within the disciplinary fields of Business and Economics. These studies share the same objectives that marketing departments of major film studios and television networks (especially overseas) have pursued since the 1980s and 1990s, with a certain

degree of optimism. The underlying idea is that audience appreciation of audiovisual products is not random, and success is not merely a matter of chance. Therefore, by studying data, it is possible to identify patterns and regularities, and even build models that can support various phases of production and distribution (Hennig-Thurau and Houston 2019). It would not be true, as is often argued in every sector of the entertainment industry, that "nobody knows anything" about what it takes for a certain product to become a hit, and that it is only right and inevitable that every strategic decision in this field is made instinctively, "by gut feeling", following intuition. In fact, a quantitative study of audiovisual texts on the one hand, and of markets and consumption on the other, would make it possible to identify factors that attract viewers. This, in turn, can "substantially [increase] the probability of success in the entertainment industry" (Hennig-Thurau and Houston 2019) or at least predict the reception that certain content will have based on how it was made, promoted, and distributed.

If the proponents of *Entertainment Science*, as it is called, are mainly business scholars and entertainment economists, it is because it was often the companies themselves that were particularly familiar with data. And they still are, usually much more so than film and media studies academics, at least. The film industry typically studies data on box office revenue and attendance, and on the number of theatres in which films are distributed, also in relation to geographical and temporal dimensions, and integrates this data with interviews and qualitative research to better understand the characteristics of those who watch films in theatres. The television industry is even more familiar with data and has at its disposal even more granular data derived from user panels employed by consumer research and market measurement companies, which provide very precise details on the demographics of viewers, such as their socio-economic class. Digitalization has led, in Italy as elsewhere, to the integration of the classic "panel based" paradigm in the measurement of television audience data with a "census" paradigm. Total television viewing is no longer solely inferred from the behaviors of the representative panel considered by the measurement company and then projected onto the entire population through statistical data expansion factors, but is also described through tracking all individual devices enabled for viewing via IP protocol that consume content through apps or browsers being surveyed, such as tablets, smartphones, smart TVs, computers, etc., both live and on-demand, at home or away. This has introduced new metrics, and a whole series of challenges to reconcile "traditional" data with that

of the new "everytime, everywhere, everydevice" digital TV (Guarnaccia 2022 offers some insights on this topic).

From much of these challenges and data, however, academics are usually excluded, with rare exceptions (e.g., Scaglioni 2021, 2022, Thurman 2021, see also Avezzù 2022 on the geographies of film and TV consumption). This exclusion hinders the communication that we would like to see, which is not only between the humanities and the hard sciences, but also that which should result from the encounter between the university and the audiovisual industry. Notably, the use of data and of tools for their (geo)visualization, for example to describe cinemagoing behaviors, seems to be more frequent – within film studies – in reference to distant decades of film history than in reference to the contemporary scenario. This is the case, among others, with certain contributions of New Cinema History and other research on national audience histories (Maltby, Walker and Walsh 2014, Treveri Gennari et al. 2020, Baptist, Noordegraaf and van Oort, 2021), or even with attempts to elaborate new tools for measuring film popularity (Sedgwick 2009). Without detracting from all these contributions and the various strands of research they represent, which we believe to be fundamental, it is inevitable to note a greater propensity to experiment with new tools and methodologies on objects of study from which a certain historical distance separates us, while applications on more contemporary objects seem rarer, not least because of a greater scarcity of data, which may seem to be paradoxical.

One field where both the relevance of data and the lack of its availability for academic research is particularly evident is that of video-on-demand (VOD) platforms. This is especially true for subscription-based platforms, which are excluded from any official survey perimeter by measurement companies in Italy and elsewhere. As a result, they do not employ any certified or publicly shared data or metrics. This is still a very loosely regulated area, where individual players, such as Netflix, are currently free to construct their own metrics and use them at will, primarily for promotional purposes, and without any possibility of third-party verification or control. Apart from their promotional and sometimes ambiguous use, video-on-demand platforms use data for content intelligence, that is, to decide what content to produce, and to feed their recommendation systems, and thus to profile content offerings according to users' tastes. Subscription-based VOD operators have often extolled such recommender systems as fundamental to their own services (Gomez-Uribe and Hunt 2015, Hallinan and Striphas 2016), and this has solicited some interest from the academic community,

but these are, of course, objects of study that are out of reach for academics, and can only be approached as black boxes. Various attempts by media scholars to "expose" or "reverse engineer" Netflix's recommendation systems – or rather, its "algorithm" – have more than anything else demonstrated the difficulty of studying something from which one is excluded, and interrogating data and data-driven logics to which one is not granted access (Pajkovic 2021, Gaw 2022).

As we have seen through this swift and non-exhaustive survey, the application of data-driven methods to the study of audiovisual products opens up a vast array of perspectives. It also requires careful reflection on the nature of data itself: what do we talk about when we talk about data? It is also perhaps in this area that scholars from the humanities can offer an important contribution by questioning and problematizing the data. What can we do with the data today, and what can we study better than before? What research can add something new to the classic media studies debates, and in what specific subfields? What new perspectives can we gain access to that were previously unavailable to us? What data and tools do we need to achieve our goals? On the other hand, what are the limitations of data-driven methods, and are there any unnecessary or misleading approaches that cannot be brought into dialogue with traditional strands of research in audiovisual studies? Furthermore, what are the media industries already doing with data, and can we learn anything from them? As academic scholars can we contribute anything useful to them?

Below, we provide a brief overview of the contributions collected in this volume.

Mirko Degli Esposti and Guglielmo Pescatore, who collaborated with us on the curation of the 13th edition of the Media Mutations conference, authored the first article in this collection. Their essay delves into the use of data-driven approaches and computational techniques in television studies, while also addressing the challenges posed by the complex open systems of TV series that interact with external factors. The article further examines the possibilities of generative AI and the use of deep multimodal algorithms to create new content.

The second contribution of this collection focuses on the automatic content analysis of dialogues of the TV series *Gomorrah*. Paola Dalla Torre and colleagues utilize a data-driven approach to uncover the correlation between

a character and their lexical signature. They employ machine learning algorithms to process the show's standardized transcripts, analyzing various features, including word presence, lexicon similarity and variety, character presence, and character identification.

In the third contribution, Valentina Re and Marica Spalletta investigate women's employment in key behind-the-scenes roles in Italian TV crime dramas broadcast from Fall 2015 to Summer 2022. The data reveals that there are stronger gender inequalities in the Italian television production industry than in the European averages for TV fiction and the Italian film sector, despite some positive trends.

Barotsi and colleagues examine the relationship between the "ecology of knowledge", data, and gender equality policies, using the CENTRIC project as a starting point. They discuss the methodological challenges in creating an open and participatory research platform on gender inequality in the audiovisual industry, which can serve as a powerful tool for promoting inclusiveness and cultural pluralism.

Andrzej Meler and colleagues analyze how modern Polish TV series attempt to challenge and reinforce stereotypes about women through their depiction of female characters. They employ a multi-layered coding approach to measure the contexts in which female characters are presented and a qualitative analysis to identify the stereotypes or anti-stereotypes presented in the series.

Marta Boni's essay delves into the potential of open-access digital tools for studying media content as a space of discourse. By integrating both quantitative and qualitative analysis, Boni argues that such an approach can reveal interesting features that may not be identifiable through qualitative analysis alone and the adaptation and testing of tactics for the specific object of study can provide deeper understanding.

Stefania Antonioni and Dom Holdaway employ data-driven methods to analyze the reception practices and discussions surrounding the Italian medical drama series *Doc – Nelle tue mani* on social media platforms. Their study examines user comments, discussions, and visual content on Instagram and YouTube, utilizing both quantitative and qualitative analyses of scraped data. In their analysis, the authors explore the limitations and challenges of data-driven approaches in such studies.

BIBLIOGRAPHY

Anantrasirichai, Nantheera and David Bull (2022). "Artificial Intelligence in the Creative Industries: A Review." *Artificial Intelligence Review* 55: 589-656.

Avezzù, Giorgio (2022). *L'Italia che guarda. Geografie del consumo audiovisivo*. Roma: Carocci.

Baptist, Vincent, Julia Noordegraaf and Thunnis van Oort (2021). *"*A Digital Toolkit to Detect Cinema Audiences of the Silent Era: Scalable Perspectives on Film Exhibition and Consumption in Amsterdam Neighbourhoods (1907-1928)." *Studies in European Cinema* 18(3): 252-73. https://doi.org/10.1080/17411548.2021.1930974.

Bednarek, Monika (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge: Cambridge University Press.

Bednarek, Monika (2020). "On the Usefulness of the Sydney Corpus of Television Dialogue as a Reference Point for Corpus Stylistic Analyses of TV Series." In *Telecinematic Stylistics*, edited by Christian Hoffman and Monika Kirner-Ludwig, 39-62. London: Bloomsbury.

Bojanowski, Piotr, Francis Bach, Ivan Laptev, Jean Ponce, Cordelia Schmid and Josef Sivic (2013). "Finding Actors and Actions in Movies." In *Proceedings of the IEEE international conference on computer vision*, 2280-87.

Bordwell, David (2006). *The Way Hollywood Tells It: Story and Style in Modern Movies*. Berkeley: University of California Press.

Bost, Xavier, Serigne Gueye, Vincent Labatut, Martha Larson, Georges Linares, Damien Malinas and Raphaël Roth (2019). "Remembering Winter was Coming: Character-Oriented Video Summaries of TV Series." *Multimedia Tools and Applications* 78: 35373-99. https://doi.org/10.1007/s11042-019-07969-4.

Chapoton, Boris, Anne-Laure Werlen and Véronique Regnier Denois (2020). "Alcohol in TV Series Popular with Teens: A Content Analysis of TV Series in France 22 Years after a Restrictive Law." *European Journal of Public Health* 30(2): 363-68. https://doi.org/10.1093/eurpub/ckz163.

Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens and Julia Noordegraaf (2021). "Archives and AI: An Overview of Current Debates and Future Perspectives." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15(1):1-15. https://doi.org/10.1145/3479010.

Cour, Timothee, Benjamin Sapp, Akash Nagle and Ben Taskar (2010). "Talking Pictures: Temporal Grouping and Dialog-Supervised Person Recognition." In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1014-21.

Cutting, James E. and Ayse Candan (2015). "Shot Durations, Shot Classes, and the Increased Pace of Popular Movies." *Projections* 9(2): 40-62. https://doi.org/10.31 67/proj.2015.090204.

De Quintana, Eugenio López and Antonio León Carpio (2021). "Artificial Intelligence for a Role Change in Television Archives: The Atresmedia–Etiqmedia Experience." *Journal of Digital Media Management* 10(2): 177-87.

Derda, Izabela (2022). "'Did you Know that David Beckham Speaks Nine Languages?': AI-Supported Production Process for Enhanced Personalization of Audio-Visual Content." *Creative Industries Journal* 1-16. https://doi.org/10.1080/17510694.2021.2025001.

Doukhan, David, Géraldine Poels, Zohra Rezgui and Jean Carrive (2018). "Describing Gender Equality in French Audiovisual Streams with a Deep Learning Approach." *VIEW Journal of European Television History and Culture* 7(14): 103-22. https://doi.org/10.25 969/mediarep/14757.

Fanchi, Mariagrazia and Matteo Tarantino (2019). "Donne e reti creative nel cinema italiano (2004-2016)." *Economia della Cultura* 29(4): 523-30. https://doi.org/10.1446/96648.

Fanchi, Mariagrazia and Matteo Tarantino (2020). "Has Digital Transformation Impacted Gender Imbalance in Italian Cinema? A Data Analysis of Creative Clusters 2004–2016". In *Digital Transformation in the Cultural and Creative Industries*, edited by Marta Massi, Marilena Vecco and Yi Lin, 162-184. London: Routledge.

Fischer, Mika, Hazım Kemal Ekenel and Rainer Stiefelhagen (2011). "Person Re-Identification in TV Series Using Robust Face Recognition and User Feedback." *Multimedia Tools and Applications* 55: 83-104. https://doi.org/10.1007/s11042-010-0603-2.

Fronzetti Colladon, Andrea and Maurizio Naldi (2019). "Predicting the Performance of TV Series through Textual and Network Analysis: The Case of *Big Bang Theory*." *PloS ONE* 14(11): e0225306. https://doi.org/10.1371/journal.pone.0225306.

Gaw, Fatima (2022). "Algorithmic Logics and the Construction of Cultural Taste of the Netflix Recommender System." *Media, Culture & Society* 44(4): 706-25. https://doi.org/10.11 77/01634437211053767.

Ghiassi, Manoochehr, David Lio and Brian Moon (2015). "Pre-Production Forecasting of Movie Revenues with a Dynamic Artificial Neural Network." *Expert Systems with Applications* 42(6): 3176-93. https://doi.org/10.1016/j.eswa.2014.11.022.

Gomez-Uribe, Carlos A. and Neil Hunt (2015). "The Netflix Recommender System: Algorithms, Business Value, and Innovation." *ACM Transactions on Management Information Systems* 6(4) art. 13: 1-19. https://doi.org/10.1145/2843948.

González-de-Garay, Beatriz, María Marcos-Ramos and Carla Portillo-Delgado (2020). "Gender Representation in Spanish Prime-Time TV Series." *Feminist Media Studies* 20(3): 414-33. https://doi.org/10.1080/14680777.2019.1593875.

Gorinski, Philip and Mirella Lapata (2015). "Movie Script Summarization as Graph-Based Scene Extraction." In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1066-76.

Greenberg, Bradley S., Rosaen Sarah F., Worrell Tracy R., Salmon Charles T. and Julie E. Volkman (2009). "A Portrait of Food and Drink in Commercial TV Series." *Health Communication* 24(4): 295-303. https://doi.org/10.1080/10410230902889233.

Guarnaccia, Fabio (ed.) (2022). *Metrix. Viaggio all'ultima frontiera delle metriche. Link. Idee per la televisione* 28.

Hallinan, Blake and Ted Striphas (2016). "Recommended for You: The Netflix Prize and the Production of Algorithmic Culture." *New Media & Society* 18(1): 117-37.

Heftberger, Adelheid (2018). *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work*. Cham: Springer.

Hennig-Thurau, Thorsten and Mark B. Houston (2019). *Entertainment Science: Data Analytics and Practical Theory for Movies, Games, Books, and Music*. Cham: Springer.

Himes, Susan M. and J. Kevin Thompson (2007). "Fat Stigmatization in Television Shows and Movies: A Content Analysis." *Obesity* 15(3): 712-18. https://doi.org/10.1038/oby.2007.635.

Holdaway, Dominic (2016). "La rete sociale del cinema di interesse culturale." In *Il cinema di Stato. Finanziamento pubblico ed economia simbolica nel cinema italiano contemporaneo*, edited by Marco Cucco and Giacomo Manzoli, 127-69. Bologna: Il Mulino.

Huang, Qingqiu, Lei Yang, Huaiyi Huang, Tong Wu and Dahua Lin (2020). "Caption-Supervised Face Recognition: Training a State-of-the-Art Face Model without Manual Annotation." In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII*, 139-55. Cham: Springer International Publishing.

Hunter, Starling (2019). "Predicting Nielsen Ratings from Pilot Episodes Scripts: A Content Analytical Approach." *SERIES – International Journal of TV Serial Narratives* 5(1): 9-21. https://doi.org/10.6092/issn.2421-454X/9079.

Jensen, Klaus Bruhn (2020). *A Handbook of Media and Communication Research: Qualitative and Quantitative Methodologies*. London: Routledge.

Lasswell, Harold Dwight (1927). *Propaganda Techniques in the World War*. New York: Knopf.

Loist, Skadi and Elizabeth Prommer (2019). "Gendered Production Culture in the German Film Industry." *Media Industries* 6(1): 95-115.

Macnamara, Jim (2018). "Content Analysis." In *Mediated Communication*, edited by Philip M. Napoli, 191-212, Berlin-Boston: De Gruyter. https://doi.org/10.1515/9783110481129-012.

Maltby, Richard, Dylan Walker and Mike Walsh (2014). "Digital Methods in New Cinema History." In *Advancing Digital Humanities*, edited by Katherine Bode and Paul Longley Arthur, 95-112. London: Palgrave Macmillan.

Mbunge, Elliot, Stephen Gbenga Fashoto and Happyson Bimha (2022). "Prediction of Box-Office Success: A Review of Trends and Machine Learning Computational Models." *International Journal of Business Intelligence and Data Mining* 20(2): 192-207. https://doi.org/10.1504/IJBIDM.2022.120825.

Moretti, Franco (2022). *Falso movimento. La svolta quantitativa nello studio della letteratura*. Milano: Nottetempo.

Nagrani, Arsha and Andrew Zisserman (2018). "From Benedict Cumberbatch to Sherlock Holmes: Character Identification in TV Series without a Script." *arXiv preprint arXiv:1801.10442*. https://doi.org/10.48550/arXiv.1801.10442.

Napoli, Philip M. (1999). "Deconstructing the Diversity Principle." *Journal of Communication* 49(4): 7-34. https://doi.org/10.1111/j.1460-2466.1999.tb02815.x.

Noordegraaf, Julia (2016). "Computational Research in Media Studies: Methodological Implications." *Kwalon* 21(1): 52-59. https://doi.org/10.5117/2016.021.001.007.

Pajkovic, Niko (2021). "Algorithms and Taste-Making: Exposing the Netflix Recommender System's Operational Logics." *Convergence: The International Journal of Research into New Media Technologies* 28(1): 214-235. https://doi.org/10.1177/13548565211014464.

Papalampidi, Pinelopi, Frank Keller, Lea Frermann and Mirella Lapata (2020). "Screenplay Summarization Using Latent Narrative Structure." *arXiv preprint arXiv:2004.12727*. https://doi.org/10.48550/arXiv.2004.12727.

Parkhi, Omkar M., Esa Rahtu and Andrew Zisserman (2015). "It's in the Bag: Stronger Supervision for Automated Face Labelling." *International Conference on Computer Vision.*

Ramos, María Marcos, Beatriz González de Garay and Carla Portillo Delgado (2019). "The Representation of Immigration in Contemporary Spanish Primetime TV Series." *Revista Latina de Comunicación Social* 74: 285-307.

Rocchi, Marta and Guglielmo Pescatore (2022). "Modeling Narrative Features in TV Series: Coding and Clustering Analysis." *Humanities and Social Sciences Communications* 9, 333. https://doi.org/10.1057/s41599-022-01352-9.

Ruffino, Marco and Paola Brembilla (2016). "Narrative Ecosystems through the Network Analysis Lens. Step One: The Production of US TV Series, between Capital and Labor Strategies." *SERIES –International Journal of TV Serial Narratives* 2(1): 55-68. https://doi.org/10.6092/issn.2421-454X/6164.

Salt, Barry (1974). "Statistical Style Analysis of Motion Pictures." *Film Quarterly* 28(1): 13-22. https://doi.org/10.2307/1211438.

Salt, Barry (1983). *Film Style and Technology: History and Analysis*. London: Starword.

Scaglioni, Massimo (2021). *La televisione nella pandemia. Intrattenimento, fiction, informazione e sport nell'anno del Covid-19. Annuario 2021*. Roma: Carocci.

Scaglioni, Massimo (ed.) (2022). *Total TV. Intrattenimento, fiction, informazione e sport fra broadcasting e streaming. Annuario 2022*. Milano: Certa.

Sedgwick, John (2009). "Measuring Film Popularity: Principles and Application." In *Digital Tools in Media Studies: Analysis and Research: An Overview*, edited by Michael Ross, Manfred Grauer, Bernd Freisleben, 43-54. Bielefeld: Transcript Verlag.

Sink, Alexander and Dana Mastro (2017). "Depictions of Gender on Primetime Television: A Quantitative Content Analysis." *Mass Communication and Society* 20(1): 3-22. https://doi.org/10.1080/15205436.2016.1212243.

Smith, John R., Dhiraj Joshi, Benoit Huet, Winston Hsu and Jozef Cota (2017). "Harnessing AI for Augmenting Creativity: Application to Movie Trailer Creation." In *Proceedings of the 25th ACM International Conference on Multimedia*, 1799-1808.

Snow, Charles P. (1959). *The Two Cultures and the Scientific Revolution*. London: Cambridge University Press.

Thurman, Neil (2021). "When a TV Channel Reinvents itself Online: Post-Broadcast Consumption and Content Change at BBC Three." *Convergence: The International Journal of Research into New Media Technologies* 27(2): 291-312. https://doi.org/10.1177/1354856520967773.

Toubia, Olivier, Garud Iyengar, Renée Bunnell and Alain Lemaire (2019). "Extracting

Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption." *Journal of Marketing Research* 56(1): 18-36.

Toubia, Olivier, Jonah Berger and Jehoshua Eliashberg (2021). "How Quantifying the Shape of Stories Predicts their Success." *Proceedings of the National Academy of Sciences* 118(26): e2011695118. https://doi.org/10.1073/pnas.2011695118.

Treveri Gennari, Daniela, Catherine O'Rawe, Danielle Hipkins, Silvia Dibeltulo and Sarah Culhane (2020). *Italian Cinema Audiences: Histories and Memories of Cinemagoing in Post-War Italy*. New York: Bloomsbury.

Tsivian, Yuri (2009). "Cinemetrics, Part of the Humanities' Cyberinfrastructure." In *Digital Tools in Media Studies: Analysis and Research: An Overview*, edited by Michael Ross, Manfred Grauer and Bernd Freisleben, 93-100. Bielefeld: Transcript Verlag.

Tukachinsky, Forster, Rebecca Riva, Dana Mastro and Marko Dragojevic (2022). "Linguistic Characteristics of Interracial Interactions on Primetime TV: A Quantitative Content Analysis." *Western Journal of Communication* (2022): 1-25. https://doi.org/10.1080/10570314.2022.2146462.

Verhoeven, Deb, Bronwyn Coate and Vejune Zemaityte (2019). "Re-Distributing Gender in the Global Film Industry: Beyond #MeToo and #MeThree." *Media Industries Journal* 6(1): 135-55. https://doi.org/10.3998/mij.15031809.0006.108.

Verhoeven, Deb, Katarzyna Musial, Stuart Palmer, Sarah Taylor, Shaukat Abidi, Vejune Zemaityte and Lachlan Simpson (2020). "Controlling for Openness in the Male-Dominated Collaborative Networks of the Global Film Industry." *PloS ONE* 15(6): e0234460. https://doi.org/10.1371/journal.pone.0234460.

A DOUBTFUL HANDSHAKE: THE PROMISING
FRONTIER OF DATA-DRIVEN FILM AND MEDIA STUDIES